Diversity Trumps Ability: Reassessment, Efficiency & Beyond

Tabio Romanski S4113306

Campus Fryslân University of Groningen

Capstone Project (CFB063A10): Bachelor Thesis

Dr. Ofer Engel

10 June 2022

Abstract

Today's challenges require collaboration, thus team composition plays an essential role. This thesis investigates the importance of diversity and ability by taking up the current "Diversity trumps Ability"-debate. An agent-based-modelling approach is used to infer the relative importance and interactions of individual ability of group members and diversity in epistemic groups and distinguish these properties from the predominant approach of focus on group assembly.

Further, this contribution adds the novel aspect of efficiency. Efficiency can play a more essential role than the optimal solution, on which the current debate is focused. Thus, quantifying the resources used to reach a given outcome presents an important goal.

I investigate the association between each of the group properties and their interaction with the performance indicators of outcome and efficiency across a variety of dimensions using plots, regression analysis and correlation coefficients. Thereby, it is found that a synergy of ability and diversity leads to good optimal solutions. The dominant factor depends on the task type. Efficiency paints a more nuanced picture with a variety of influential factors, which leads to ability and diversity being either harmful or beneficial. These findings suggest potential for diverse groups with highly able individuals under certain conditions. Further research can investigate the potential and nuances both properties have empirically, particularly concerning efficiency.

Keywords: diversity trumps ability, diversity, ability, efficiency, epistemic groups, agent-based-modelling

Table of Contents

Selecting team members for high stakes positions, not on their merit, but by random chance sounds irrational at best. But this strategy has been shown to improve the team's performance under certain conditions (Hong & Page, 2004). One might ask: how is this possible and what are those conditions? The driver of the outperformance is functional diversity. In other words, adding as many different skills and abilities as possible to a team will boost the performance of several top experts in one field, if the task is complex and the development of future results is hardly foreseeable, which are two of the conditions (Hong & Page, 2004). Arguably, the greatest challenges, such as achieving the Sustainable Development Goals, are complex and the future is inherently difficult to predict, making the findings highly likely to be relevant.

Others have debated the operationalisation of these conditions (Grim et al., 2018). Thus, due to the following debates and multiple lines of research, this paper investigates the conditions necessary through a clear distinction between the group properties of individual ability and diversity, including a widened understanding of diversity introduced by Singer (2018). Additionally, I add the novel aspect of efficiency of diverse groups while past research focuses on a group's final outcomes (what a group can achieve eventually) regardless of any procedural insights i.e. how a group reaches the final outcome.

Past research has focused on contrasting groups with high-achieving individuals, which tend to be similar to each other, and randomly assembled groups, which tend to be comparatively more diverse, using computer simulations (Hong & Page, 2004). This research has led to controversies around the explanatory value of diversity (Thompson, 2014). Hong and Page (2004) have shown that the randomly assembled groups outperform the groups of individually best-performing agents on average. They inferred that this difference can be explained by the diversity of these groups. However, it has been argued that the randomly composed ones are more successful due to their randomness[1] rather than their diversity[2] (Thompson, 2014). This argument has been seriously challenged by Singer (2018) on the basis of an alternative notion of diversity, highlighting and demonstrating the importance of a theoretical understanding of the concept.

Additionally, the general applicability of the diversity trumps ability (DTA) argument on the basis of the original simulation has been critiqued. Grim et al. (2018) have argued that entirely random and unforeseeable tasks do not represent realistic scenarios and have shown that expert groups tend to outperform randomly assembled groups (often called random groups) on correlated landscapes on the basis of Hong and Page's (2004) definition of diversity combined with variations of agents and group dynamics.

In summary, there are debates on several topics surrounding the model introduced by Hong and Page (2004). On the one hand, the importance of the process of assembly vs diversity as a property of the group is discussed (Hong & Page, 2004; Thompson, 2014; Singer, 2018). These two are inherently interlinked as groups can be composed according to desired level of diversity (Singer, 2018; Thompson, 2014) or according to individual ability, which leads to less diversity (Hong & Page, 2004). On the other hand, the applicability of results along different dimensions regarding the task and group dynamics as well as properties of agents is questioned (Grim et al., 2018; Holman et al., 2018). Both research strains are partly interconnected yet not treated as such. Here, I aim to combine, order and unify these themes.

Besides, several factors such as time-efficiency of groups and cost of communication between agents have not been explored but merely suggested (Hong & Page, 2001, 2004) and

---

[1] Randomness refers to the process of assembling a group, it means that members of the group have been chosen at random from the pool of all agents

[2] This is a property of the group, which can be acquired irrespective of the process of assembly

remain relevant in terms of applicability of the theoretical findings. Consequently, this paper explores the efficiency groups of agents to expand the understanding of success by adding procedural insights. While incorporating efficiency does not change the model with regards to communication in any way and thus introduces no cost thereof, it can be insightful regarding the potential impacts of such costs by quantifying the process groups go through to reach their final result.

Overall, these gaps and disconnections in the literature can be conceptualised using three factors. Firstly, the process of assembly has played a major role in previous research starting with Hong and Page (2004) and continued by Thompson (2014) and Singer (2018) arguing over the influence of the process. Secondly, this aspect has been mixed with properties of the group. Here, diversity has played a major role (including the introduction of two different notions of diversity), however, individual ability as well in the form of the reference group, usually the group type of individually best-performing agents (Hong & Page, 2004). Thirdly, performance or more precisely the impact of the former two on performance is the final factor and includes outcome (the answer to ‘How good of a result does a group reach eventually?’) but should also entail efficiency (resources used per output). Given the confusion around the first two factors and the lack of efficiency in earlier literature, I propose the following research questions to illuminate these gaps and unclarities by comparing the four different types of groups (randomly assembled, maximally diverse (one per measure) and individually best-performing) and distinguishing it from the group properties of average individual ability and diversity.

1. How does the type of group assembly (random, individually best-performing or maximally diverse) affect performance (outcome and efficiency)?
2. What is the association between the group properties and performance?

a) What is the association between group diversity and performance?

b) What is the association between average individual ability and performance?

c) What is the association between the group properties as well as their interactions and performance?

Moreover, there are several dimensions which have been altered in the literature with decisive differences in the results. These include the type of task, particularly its predictability (Grim et al., 2018; Holman, 2018), the potential variance of functional diversity within one agent and the mode of deliberation i.e. the way agents collaborate in a given group (Grim et al., 2018; Holman, 2018, Singer, 2018). These factors are considered and varied for each research question. The factors are already explored in the literature however, neither explicitly in relation to both group properties but rather as the mix of group composition and properties, nor with the inclusion of the second diversity measure by Singer (2018), hence the synthesis with the addition of efficiency here.

The remainder of the thesis is structured in five sections. First, a literature review of the relevant contributions is conducted. Thereby, special emphasis is put on the contributions which have conducted similar experiments to justify the research questions based on the identified gap. Additionally, the three dimensions are motivated and elaborated on. Second, the model and the analytical tools to evaluate the results are introduced in detail. Thirdly, the results are reported, structured along the research questions above. These findings and their practical interpretations are discussed in light of the literature review and other related approaches to highlight the weaknesses of the model itself and offer potential solutions. Finally, I conclude with the overall finding that carefulness should prevail when using the DTA argument regarding outcome and efficiency depending on the context. Thus, further specific and empirical research is needed to fully understand the interplay of diversity and ability.

Overall, I aim at providing clarity on the core aspects of diversity and ability which have been incorporated in various earlier contributions yet not structured and combined. Hence, this thesis focuses on these aspects and does not explore other related approaches in-depth although some are discussed to contrast the models and open future avenues to remedy potential shortcomings of the model here. Neither does this thesis analyse the implications for domains, which have taken the findings by Hong and Page (2004) among others and have drawn their own conclusions e.g. in the field of epistemic democracy (Landemore, 2021). Besides, the initial contribution contains the computational experiment as well as a mathematical proof for the DTA argument (Hong & Page, 2004). This thesis elaborates on the former for the sake of conciseness as both have partially distinct assumptions (Thompson, 2004) and would therefore require substantially different inquiry. Hence, the primary object of inquiry is the model and the following simulations originally introduced by Hong and Page (2004) and the literature directly building up on it and modifying the simulation.

Literature review

This section reviews the most relevant literature on the DTA debate focusing on research that uses and advances computational modelling. Thereby, the existing literature is presented, synthesised and subsequently gaps are identified leading to the formulation of the research questions. Mainly approaches from various papers are used and combined with each other to offer a comprehensive view on the debate and findings. Additionally, a procedural view on the models is introduced inspired by research from related fields. Hence, the literature review presents a brief yet comprehensive overview of computational models following Hong and Page (2004).

The initial model was introduced by Hong and Page (2004) which followed an earlier paper with a similar approach (Hong & Page, 2001). They have found that a randomly assembled group of varying size outperforms a group of the individually best-performing agents on average on a given task (understood as epistemic search). These findings build on several assumptions. Three of these assumptions are crucial for the interpretation of the results and are thus introduced here.

1. The tasks are complex. The criterion for a task to be complex is that the task cannot be solved by any individual perfectly by themselves. Such a task is represented by a value function in which each position holds a value (see methodology for details), and the agent is required to find the highest value in the landscape. These tasks are not predictable meaning that from the progress made by an agent the future cannot be predicted as the values are drawn randomly. Additionally, the length can vary.
2. Agents are unique and can be represented as an ordered set of numbers (usually three numbers) representing their problem-solving heuristics, where each number is conceived as one heuristic. These numbers lie between one and usually twelve or twenty. The upper bound is called maximum heuristic and determines the heuristic space. They indicate which positions of the value function an agent can explore to find a higher value e.g. 20 would mean to be able to look at the position 20 positions ahead from their current one.
3. Value function and agents together provide the basis for testing randomly assembled the group type and the group type of best-performers. This is done by either picking agents randomly from the pool of all agents or having them explore a value function individually

and pick the ones with the best expected result on a value function. Group size as well as the dynamics within a group (deliberation dynamics/modes of collaboration) vary as well.

Overall, Hong and Page find qualitatively similar results when varying factors such as deliberation dynamics, group sizes, the length of the value function and maximum heuristics. They attribute the overall superiority of randomly assembled groups to the diversity of this group type, which is relatively higher according to their diversity measure than the group type of best-performers.

This diversity measure is called HP-diversity (Singer, 2019). It conceptualises diversity as no overlap on the same position of the heuristics. Thus, two agents with the heuristics {1,2,3} and {3,1,2} are maximally diverse as the ordered pairwise comparison of heuristics (position one of agent 1 vs position one of agent one etc.) yields no overlap at any position although the same three numbers are present in the agents' heuristics (Hong and Page, 2004).

Criticism and Extensions of the Original Model

The criticism can be divided into two main directions. On the one hand, criticism is directed to the attribution of diversity as a group property and the suitability of the proposed measure of diversity (Singer, 2018; Thompson, 2014). On the other hand, the fact that best-performing agents only perform well on this one very specific landscape and thus cannot be understood as experts with transferable expertise, has been criticised. Changes to the model influencing the transferability of expertise highlight the importance of other dimensions such as maximum heuristic as well as deliberate dynamics (Grim et al., 2018; Holman et al., 2018). Hence, the review of criticism is structured along these two main themes.

Randomness vs Diversity

Thompson (2014) has pointed out that the primary driver of the results found by Hong and Page (2004) is not the diversity but rather the process of assembly, which is picking members randomly from the pool of all possible candidates. As evidence, she shows that some maximally HP-diverse groups perform worse than a randomly composed group.

Responding to the critique, Singer (2018) shows that there is a strong correlation between the diversity of a group and their performance on more than a few select groups, which he assumes is how Thompson (2004) arrived at the conclusion that randomness is the driver of the results. Additionally, Singer (2019) finds that maximally HP-diverse groups outperform random ones on average and that there are highly diverse groups, assembled to be diverse, which almost always outperform any randomly assembled group. Furthermore, several theoretical counter arguments are offered. These explain why randomness cannot be the decisive factor driving the final outcome reached by a group but rather a means to an end.

Subsequently, to offer an explanation for the maximally HP-diverse groups used in Thompson (2014), which were much worse than the randomly assembled ones and the average outcome Singer (2019) introduces Coverage Diversity (C-diversity). It quantifies the relative number of all possible heuristics present in a given group. The groups used by Thompson have a very low C-diversity and simply the same heuristics in different positions of the ordered sets. Hence, these groups have a low C-diversity but high HP-diversity. The results strongly indicate that C-diversity is an even stronger predictor of the final outcomes reached by a given group of agents. This leads to the conclusion that order seems to be less relevant than full coverage (Singer, 2019) as order is prominent in Hong and Pages' measure of diversity (Hong & Page, 2004). Furthermore, it is suspected that an even spread of extreme heuristics may be relevant to not miss a peak nor be stuck at a low. Thus, the main contribution of Singer (2019) is certainly highlighting the importance of various understandings of diversity and the effective measure of C-diversity in particular.

Critique on Expertise and its Consequences

Besides the interpretation of the results by Hong and Page (2004) also the design and assumptions of the model have been criticised. Grim et al. (2018) have pointed to the problematic notion of 'experts' or high-ability problem solvers. The notion of genuine expertise "requires being able to perform well on many problems of the same type, not just on a single problem" (Grim et al., 2018, p. 99). This criterion is often thought of as the minimal criterion for an expert and has also been used in related literature (Holman et al., 2018). The literature on the DTA argument has coined these truly expert agents to be holders of transportable or transferable expertise (Sakai, 2020).

Grim et al. (2018) proceed with a slightly modified model: they introduce smoothed landscapes in which adjacent values are correlated. The findings indicate that experts indeed fulfil the minimal criterion, and that groups of experts outperform randomly assembled groups with increasing smoothing factor (Grim et al., 2018; Holman et al., 2018). Additionally, both aspects appear at about the same smoothness together with clear patterns in the heuristics of the experts (Grim et al., 2018).

Moreover, the findings suggest that a change in maximum heuristics can yield results in favour of DTA. Grim et al. (2018, p. 114) summarise their interpretation as follows: "Groups of the best-performing are better for a wide range of smoothness but only where the available conceptual resources are relatively limited. With a wider pool of conceptual resources, a diverse group will do better even on problems of that same character".

Furthermore, they have tested both modes of collaboration/deliberation dynamics and mixed groups of individually best-performing and randomly selected agents. Both aspects have their relevance as mixed groups tend to occur in practical debates and it is informative from a normative point of view whether such groups perform better. Also, both deliberation dynamics may be practical. For the one coined relay or sequential (Hong & Page, 2004), Weymark (2015) has argued that it is highly stylized yet contains key-features of debates (reasoning, proposals, formalised responses and repetition once moved on) as each agent performs until a local optimum is reached to pass the result onto the next agent until the group's local optimum is reached. The tournament mode (Hong & Page, 2004) similarly contains some features which resemble competition in the sense that the best proposal of all agents is selected after they all have reached their local optimum. This best proposal is then the basis for further development with simultaneous efforts of all involved. Hence, combining all these aspects were suspected to yield new insights

The findings on these two modelling variations indicate general advantages of diverse groups in the non-sequential/tournament mode as well as mixed groups consisting of experts and non-experts partially depending on group size and smoothness (Grim et al., 2018). Holman et al. (2018) conducts similar experiments and confirms the patterns concerning smoothness, mixed groups, group sizes and heuristic space which tend to influence the relative importance of diversity.

Research Gaps

From these two broad debates outlined above some patterns can be derived. Firstly, there is the question of how to assemble any given group and the following expectations on the outcome: picking members randomly, choosing the individually best-performing, or assembling a group such that it is maximally diverse, if so, according to which diversity measure. Secondly, three out of four group types are assembled according to the criteria of individual ability or the group diversity, thus they keep playing a role. Thirdly, there is the debate on the diversity measure and its association with the outcome relatively independent from individual ability. I argue that these three aspects are not clearly and thoroughly explored in the literature due to the mix of group properties with processes of assembly. This unclarity informs the formulation of my research questions.

While Grim et al. (2018) and Holman et al. (2018) focus on assembling the groups according to certain criteria, Singer (2018) primarily focuses on the associations between diversity and outcome. However, the former neglects the systematic comparison between different diversity measures. Yet, the latter does not pay much attention to individual ability as a property. Hence, there are still gaps in the literature when groups are assembled according to certain criteria particularly regarding both diversity measures present in the literature. Additionally, there are gaps regarding the interplay of individual ability and diversity and their joint impact on the performance beyond mixing groups of experts with non-experts. Individual ability is often only a criterion when assembling groups but not as group property for e.g. randomly assembled groups while this is the case for diversity in Singer (2018). Therefore, this thesis strives to fill these gaps by comparing the four different types of groups (randomly assembled, maximally diverse (one per measure) and individually best-performing) and distinguishing it from the group properties of average individual ability and diversity.

Furthermore, the current literature is almost exclusively focused on the outcome or the effectiveness of groups. In part, the group composition beyond the diversity measures and its impact on the dynamics of arriving at a given point has been considered (Grim et al., 2018;

Singer, 2018). Others have taken up the DTA-debate to develop and empirically investigate a notion of efficiency, which is not clearly distinguishable from effectiveness. They understand efficiency as "being better" or solving a problem more often than others (Morand-Ferron & Quinn, 2011) as opposed to clearly defining it as solving it with less resources. However, an aggregate measurement of efficiency and thus procedural aspects of the simulation do not play a role in the current academic debate surrounding the model.

Nevertheless, in other fields it is of great importance. Moreover, some of these could benefit from the insights generated by the agent-based modelling (ABM) approach. For instance, while in political philosophy the epistemic superiority of diverse groups has often been discussed (Landemore, 2021; Holman et al., 2018) other, highly related fields tend to pay attention to the efficiency of political organisations or even political systems such as comparative politics (Lijphart, 2012). Furthermore, economic examples are often used to illustrate the value of superior decision-making (Hong & Page, 2001; Weymark, 2015). These are decisions where efficiency plays an obvious role similarly to any decision taken within firms who may not need to arrive at the best possible outcome but need to reach a satisfactory outcome in an efficient manner. This is further illustrated by the theory of bounded rationality, where satisficing plays a major role (Katsikopoulos, 2014) and has been found empirically in epistemic searches (Agosto, 2002). The concept describes the phenomenon of humans and organisations seeking satisfactory solutions due to costliness of the search and the boundaries of their own cognition. Thus, investigating the output per input (efficiency) of groups is more relevant than investigating the optimal solution, which can be reached by a group (final outcome). Additionally, other findings suggest that diversity of agents can accelerate problem solving (Clearwater et al., 1991). Therefore, I propose to investigate the efficiency of the groups. The investigation of efficiency can be done in terms of the number of checks (cognitive task of an individual agent) as well as in terms of moves (agreeing on intermediate solutions) as both can indicate distinct aspects of the process requiring resources.

Overall, systematically including the different ways the groups can be composed (that means including the group types/different ways to assemble a group) as well as both group properties leads to the following two main questions below. Additionally, I introduce the notion of efficiency which is explored for each research question the same way the final outcome is.

1. How does the type of group assembly (random, individually best-performing or maximally diverse) affect performance (outcome and efficiency)?
2. What is the association between the group properties and performance?
   a) What is the association between group diversity and performance?
   b) What is the association between average individual ability and performance?
   c) What is the association between the group properties as well as their interactions and performance?

Beyond that, several dimensions which have been shown to be influential are varied at each state. These include correlation of adjacent positions in a value function, the deliberation dynamics and maximum heuristic (Grim et al., 2018; Holman et al., 2018; Singer, 2018). Altogether, answering these questions gives a comprehensive overview of current research with filling minor gaps concerning the different types of groups, structures the debate on the interplay of group properties and adds an understanding of efficiency.

Methodology

The paper follows the influential contributions of Hong and Page (2001, 2004) as well as Grim et al. (2018)/Holman et al. (2018) and Singer (2018). Their approach falls under ABM, which has a number of insight- and hypothesis generating potentials (Grim et al., 2013) and finds application in a variety of fields with a growing number of publications (Niazi & Hussain, 2011). This particular model allows wide-ranging insights regarding group performance in relation to diversity due to its general nature and has sparked wide-ranging controversies across many domains (Bove & Elia, 2017; Grim et al., 2020; Landemore, 2021). Hence, by extending the simulation approach introduced by Hong and Page (2004), I contribute not only to a better understanding of the performance of simulated groups but also to the surrounding discussions regarding the DTA argument (Bove & Elia, 2017; Grim et al., 2020; Landemore, 2021). The following section describes the setup of the simulation first including value functions, agents, the modes of collaboration, formal introduction of the diversity measures and secondly, elaborates on the analytical tools to then justify the overall approach of computational simulations.

The Simulation

The building blocks of the simulation are value functions and agents. A value function is an ordered sequence of length 2000 with values drawn from a uniform distribution of [0,100] (Hong &Page, 2004) or alternatively correlated (Grim et al., 2018; Holman et al., 2018). The values associated with each point can be thought of as epistemic payoffs simulating the goodness of a solution (Singer, 2018).

Agents are defined by their heuristics: an ordered set of integers of size three with numbers between one and twelve or twenty (both are considered as they have been shown to be relevant (Grim et al., 2018; Holman et al., 2018; Singer, 2018)). Each agent checks the value of a value function at the position as many steps ahead as their current heuristic element allows to explore the function and solve the problem of finding the highest value. If the checked position has a higher value, it moves there and applies the next element in their collection. If this is not the case, the process is repeated with the next element until a local optimum is reached i.e. no beneficial move can be performed (Hong & Page, 2004). For example, if an agent a={1; 2; 3} and has a starting position of index i=1, it would first check position i=2 (1+1), if this does not have a higher value than its current position, it will apply 2 and thus check position i=3. If this move would not be beneficial and neither would be the next one, the agent has reached its local optimum as it cannot improve despite trying all its heuristics. If an agent reaches the end of the value function it checks the values from the first index onwards again as if the value function is a circle in which the last element is linked with the first (Hong & Page, 2004). Hence, an agent's heuristics determine its problem-solving ability given a value function.

The performance is evaluated by the highest reached number (outcome) (Hong & Page, 2004) as well as the number of steps performed by an agent or a group[3] and the number checks[4] it had to perform in relation to the outcome (efficiency) averaged over all possible starting positions in a value function. Thus, efficiency is defined as the expected final outcome divided by the expected number of checks or moves. The relative number allows insights regardless of the outcome itself and enables quantification of the number of moves/checks it takes to reach any given outcome, not only the specific one reached, which would be the case if absolute numbers were used. Besides, best-performing agents (the ones with the highest outcome) can be filtered out and grouped together by averaging their performance over all possible starting points of a value function, the common measurement for expected outcome (Hong & Page, 2004). Similarly, average individual ability as the mean of expected outcome of each group member can be computed as a group property to inform research question two.

[3] Changes of position of an agent or a group
[4] Comparisons whether a candidate position has a higher value

Groups of Agents exploring distinct Landscapes

In the following, groups of experts, random, or maximally diverse groups can be composed and explore the value function once again as a group to measure their outcome and efficiency. The group type of experts are the groups of agents composed of the individually best-performing agents while the maximally diverse group type are the groups with the highest diversity score according to a given metric, whereas the random group type are the groups which are randomly composed from the pool of all possible agents (Hong & Page, 2004; Grim et al., 2018, Singer, 2018; Thompson, 2014). The sizes of groups vary between ten and 20 agents as in Hong and Page (2004) however, this paper only considers groups of ten agents for the sake of simplicity. Finally, a group's performance can be measured the same way as for individual agents (see above). Comparing the performances of the described group types enables one to answer research question one and lets us understand how the performances relate to different processes used to assemble a given group.

Groups can collaborate/deliberate in different modes. One mode of collaboration can be outlined as follows: an agent explores a given value function until a local optimum is reached and then the following agent does the same. This process is repeated for all agents and then repeated beginning with the first one until no further progress can be made by any agent i.e. the group's local optimum is reached (Hong & Page, 2004). This mode is called sequential by Hong and Page (2004) or relay in other (Sakai, 2020) and yields similar results in their experiments as what they call the simultaneous mode of collaboration (Hong & Page, 2004) or tournament elsewhere (Sakai, 2020). At each position the best agent's outcome is picked after all agents have reached their local optima. The process of finding the best agent is repeated from the new starting point until the local optimum for this group is reached. Both collaborative modes are considered for this paper as they are likely to have an impact on outcome (Grim et al., 2018; Holman et al., 2018) and efficiency due to their procedural impact as well as for theoretical relevance (see literature review).

Diversity and Performance on different Landscapes

Each group can be characterised based on a diversity score according to different measures. Such measures are proposed by Hong and Page (2004) (HP-diversity) as well as Singer (2018) (C-diversity). HP-diversity of two heuristics ϕ^a , ϕ^b of the same size k is defined as $\frac{k - \sum_{i=1}^{k} \delta(\phi^a_i, \phi^b_i)}{k}$ where $\delta(\phi^a_i, \phi^b_i) = 1$ if $\phi^a_i = \phi^b_i$ else 0. The diversity of each possible pair of agents in a group is averaged for a given group to obtain a group score, which lies between 0 and 1. The latter quantifies how many heuristics n of all possible heuristics n_{max} are present in a given group: $\frac{n}{n_{max}}$. These diversity measures allow one to give each type of group (maximally diverse, random, and expert) a diversity score and investigate whether their success is associated with differences of diversity.

The different measurements can be used to answer research question one and two by comparing maximally diverse group types of each diversity measure and their respective expected outcome as well as efficiency with the ones of the random type and group type of individually best-performing agents. Maximally diverse groups are constructed by randomly composing a group of a given size and exchanging one agent at a time until the group is maximally diverse. While the process is inefficient, it avoids a deterministic composition with potentially overlooked characteristics. Subsequently, for research question one, the performances of each type of group can be compared. Alternatively, to answer research question two the performances can be related to the diversity scores for each group (including random and best-performing as well as

HP-diversity for maximally C-diverse group).[5] Thus, through the quantification of diversity and creating maximally diverse groups, the understanding of the different kinds of diversity and their association with performance is enhanced.

Finally, the same procedures are applied to investigate the outcomes on correlated landscapes. Correlated landscapes are problems on which values of consecutive positions are correlated (autocorrelation) (Grim et al., 2018). In this paper correlated landscapes are created by taking the average of neighbouring values of the current value and the current value itself (including for instance the last one and the second one for the average at position one) and replacing the current value with that average. Depending on the desired degree of correlation the windows from which the averages are obtained can be wider or smaller and the number of times this process is repeated over the full length of the value function. For this thesis the directly neighbouring values are used, and the process is repeated ten times. This approach has the advantage that it is more variable and cannot lead to a disruption of the pattern and the end of the value function. This makes it more suitable than the interpolation approach used elsewhere at the expense of lacking the so-called smoothness factor as presented in Grim et al. (2018) and Holman et al. (2018).

Analytical Tools

The performances of different group types and importance of diversity can be compared using various statistical tests. The performance of different group types can be compared using regression analysis to quantify the difference between the average performance of the types of groups on random and correlated landscapes to answer the first research question. Thereby, the dependent variable is the performance indicator and the independent variables the individual group type (dummified). This analysis allows us to compare more than two group types.

[5] See Analytical Tools section for further details on statistical measures used

Additionally, that difference is quantified in the form of the coefficient. The group type of reference is the group type of individually best-performing agents since it is most illustrative to investigate the claim of DTA.

Similarly, a group's diversity scores (Singer, 2018) and average individual ability (outcome) can be correlated with a performance indicator and thus Pearson product-moment correlation coefficients are indicators of the strength of their linear relationship to answer the first two sub-questions of the second research question. Additionally, the overall relationship between a measurement of diversity/individual ability and of performance can be investigated using regression analysis for further details on the fit. Here, the independent variable(s) are the group properties, and the performance indicators are the dependent ones (one regression per performance indicator). Therefore, both approaches are used in this paper, the former due to the comparability with earlier literature and the latter for its additional properties. Beyond that, visual analysis of the relationships through plotting supports the conclusion drawn. Both tests inform research question two.

The data is obtained by gathering performances for the different types of groups with varying settings. These settings include 1) both modes of collaboration, 2) correlated and uncorrelated landscapes and 3) maximum heuristics of twelve and 20. This process is repeated 50 times with a new value function each time and serves as the basis for the analysis and testing. The simulation is implemented in Python 3.9 (van Rossum & Drake, 2009) using NumPy (Harris et al., 2020) and random seeds for reproducibility of the results.[6]

[6] The Jupyter notebook including the code for the simulation and analysis as well as datasets used can be found here: https://doi.org/10.17605/OSF.IO/4YPK3

Advantages of the Simulation

The advantage of using a simulation as described above is that certain properties such as functional diversity are guaranteed by definition and the complex problem to solve (finding the maximum value in a function using heuristics) can be defined as well. Complex problems are opposed to trivial problems which could be solved by any agent (Hong & Page, 2004). By using random landscapes, a complex task without predictability based on current achievements can be simulated. This makes experts only experts on one specific problem – this one value function (and the ones with minor differences at individual positions). Contrastingly, a correlated landscape allows for predictions as well as experts which perform well beyond one specific landscape (landscapes with similar degrees of correlation) (Grim et al., 2018). Thus, discussing both scenarios allows one to investigate the performance of certain groups and metrics in a highly controlled environment with meaning for various complex tasks.

Similarly discussing two different scenarios of collaboration and maximum heuristics regarding all research questions allows for greater generalizability. While the relay mode does not require any direct communication between agents the second approach of selecting the best is more flexible. The former corresponds to problems in which one problem-solving entity takes the input of another one, which finds some applicability in the real world (Weymark, 2015). However, within teams the latter mode seems to be more plausible as team members are likely to propose several solutions with the best one being selected. The same applies to market dynamics.

Moreover, grasping the influence of and trends associated with the maximum number of available and relevant skills informs how different conditions change the relevance of diversity and ability. Thus, by comparing both deliberation dynamics, types of landscapes and different maximum heuristics, a greater range of real-world scenarios is covered by explicitly modelling these, it is very resource efficient and it has been argued that such simulations provide a relevant tool of hypothesis generation for empirical research (Sakai, 2020). While there are many advantages of this simulation and ABM in general (Grim et al., 2013), there are also shortcomings, on which I elaborate in the discussion section.

Results

The following section describes the research results. It is structured by the above-mentioned research questions. Each is briefly described and connected to the presented overviews (tables and figures). Subsequently, the main patterns related to each research question are highlighted and statistically interpreted based on the overviews.

1 How does the type of group assembly affect performance (outcome and efficiency)?

The first sub-question aims at a comparison of the group types. This comparison is done in terms of outcome, number of checks as well as moves needed to reach the final outcome. For each measurement, averages from all possible starting points along the value function in 50 trials are used with two different modes of collaboration on correlated and uncorrelated value functions. Table 1 refers to trials performed on agents whose heuristic spaces are between one and twelve whereas Table 2 shows the results for agents with heuristics between one and 20. Tables 3 and 4 are structured in the same way and report results on correlated landscapes. The analysis is performed using linear regression whereby the reference is the group type composed of the individually best-performing agents.

Overall, the results indicate that each group type performs significantly differently than the individually best-performing agents (reference) in most regards. Besides, the regressions fit the variability of the data well indicated by the high values of R^2. These are consistently about 0.5 or higher in Table 1 (except for regressing checks) and Table 2 (except checks non-sequential) and indicate an overall better fit with an enlarged pool of heuristics. Beyond that, the coefficients and intercepts are usually greater for the groups with agents composed from a greater number of possible heuristics. These main patterns are indicative for the explanatory power of the regressions as well as the importance of potential diversity which can arise from a greater pool of heuristics.

The results concerning the final outcome on uncorrelated landscapes indicate several patterns. Firstly, the outcomes for both trials are overall higher in the non-sequential mode of collaboration. Secondly, when agents have been composed from a larger pool of possible heuristics they perform better as a group. Thirdly, the best-performing group type regarding the final outcome is the maximally C-diverse, followed by the maximally HP-diverse and the randomly composed group. However, the latter difference from the reference group type is larger than between each of the three group types with the smallest difference between the two maximally diverse group types. These patterns reveal the importance of modes of collaboration, heuristics spaces and diversity for the case of the final outcome.

The coefficients for the relative number of checks is mostly negative and otherwise almost indistinguishable from zero, with amplified results for the sequential mode of collaboration and enlarged heuristic space. The maximally C-diverse group type needs the least checks followed by the random and then the maximally HP-diverse group type. Furthermore, the R^2 values are rather low compared to the other regression which could be rooted in the weak relationships/their reversal with enlarged heuristic space as can be seen particularly for the sequential mode of collaboration. Hence, the efficiency of checks increases with enlarged heuristic space particularly for the more diverse type of groups.

Finally, the relative number of moves reveals distinct results. While the overall efficiency concerning moves is positively associated with maximally diverse and randomly composed group types for the non-sequential mode, the opposite is the case for the sequential mode. These trends appear to be amplified for enlarged heuristic space. Furthermore, in the latter, the maximally C-diverse group type is once again the most efficient followed by the maximally HP-diverse then randomly composed group type. For the non-sequential mode of collaboration this trend is not as apparent but could potentially be reversed. Thus, there are major differences between the modes of collaboration next to the parallels found between the two different trials.

However, these results need to be read carefully. Testing the assumptions using visual analysis has revealed several issues. Firstly, the dependent variable is seldomly normally distributed, especially regarding the final outcome and partly also for the relative number of moves with strong skewness and deviations from the normal line in the probability plots. Similarly, the variance of the residuals is unequal for these two performance indicators: the variance is significantly greater for the group type of individually best-performing agents than for the other group types.[7] Thus, the results may not be entirely reliable, yet due to the high dependence of the specific configurations of heuristic space and specific diversity measures, it is reasonable to infer some relevant tendencies from the results instead of relying on specific numerical differences.

On correlated landscapes, there are mixed results regarding the influence of group assembly on outcome, number of checks and number of moves. The results of the groups with a heuristic space between one and twelve indicate significance only for number of moves and checks in the non-sequential mode of collaboration and to a limited extent also for moves in the

[7] Please consult the notebook (https://osf.io/4ypk3/) for further details on the assumption testing, plots are included there, see Appendix C for further description

sequential mode, where the diverse group types are less efficient, similar to the results found for the first research question. This shows that the group type of individually best-performing agents performs vastly indistinguishable from the other ones with few notable exceptions.

However, with increased heuristic space the results become amplified here too as seen in Table 4. There, the number of checks are higher for the maximally diverse and random groups in the non-sequential mode of collaboration, but lower in the sequential one. Additionally, the maximally diverse group types have a higher outcome than the group type of individually best-performing agents in the non-sequential mode of collaboration. The relationship of number of moves with the maximally diverse and random group type is rather positive or unclear, although the trend seems to be in the direction of more efficiency with increased heuristic space in the sequential mode. This indicates that diverse groups can be more efficient when it comes to the number of checks, but less or equally efficient concerning the number of moves, and achieve better results particularly in tournament-like deliberation dynamics and with enlarged heuristic space.

Table 1

Regression analysis of the different types of groups with maximum heuristic of twelve on uncorrelated landscapes. Each variable is a dummy variable for each type of group. The reference group type (intercept) is the group type of individually best-performing agents. The C coefficients are the coefficients for maximally C-diverse dummy variables similar to the HP coefficient, which is the coefficient for the maximally HP-diverse group type dummy variables. The columns represent the different performance indicators coupled with a specific mode of collaboration. Standard errors are reported in parentheses.

	outcome, sequential	outcome, non-sequential	checks, sequential	checks, non-sequential	moves, sequential	moves, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)
C	2.172^{***}	1.243^{***}	-0.019^{***}	0.005	-5.143^{***}	0.388^{***}
	(0.145)	(0.103)	(0.003)	(0.003)	(0.332)	(0.021)
HP	2.121^{***}	1.318^{***}	-0.012^{***}	0.011^{***}	-5.090^{***}	0.387^{***}
	(0.145)	(0.103)	(0.003)	(0.003)	(0.332)	(0.021)
random	1.927^{***}	1.132^{***}	-0.016^{***}	0.008^{***}	-4.558^{***}	0.384^{***}
	(0.145)	(0.103)	(0.003)	(0.003)	(0.332)	(0.021)
Intercept (best)	93.067^{***}	95.485^{***}	1.542^{***}	1.164^{***}	44.373^{***}	6.401^{***}
	(0.102)	(0.073)	(0.002)	(0.002)	(0.235)	(0.015)
Observations	200	200	200	200	200	200
R^2	0.614	0.528	0.185	0.072	0.631	0.721
Adjusted R^2	0.608	0.521	0.172	0.058	0.625	0.717
Residual Std. Error	0.723(df = 196)	0.513(df = 196)	0.015(df = 196)	0.015(df = 196)	1.659(df = 196)	0.105(df = 196)
F Statistic	103.863^{***} (df = 3.0; 196.0)	73.064^{***} (df = 3.0; 196.0)	14.811^{***} (df = 3.0; 196.0)	5.101^{***} (df = 3.0; 196.0)	111.659^{***} (df = 3.0; 196.0)	169.192^{***} (df = 3.0; 196.0)
Note:	Group size: 10; maximum heuristic: 12; *p<0.1; **p<0.05; ***p<0.01					

Table 2

Regression analysis of the different types of groups with maximum heuristic of twenty on uncorrelated landscapes. Each variable is a dummy variable for each type of group. The reference group type (intercept) is the group type of individually best-performing agents. The C *coefficients are the coefficients for maximally C-diverse dummy variables similar to the HP coefficient, which is the coefficient for the maximally HP-diverse group type dummy variables. The columns represent the different performance indicators coupled with a specific mode of collaboration. Standard errors are reported in parentheses.*

	outcome, sequential	outcome, non-sequential	checks, sequential	checks, non-sequential	moves, sequential	moves, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)
C	3.190***	1.764***	-0.058***	-0.006**	-8.192***	0.485***
	(0.164)	(0.110)	(0.004)	(0.003)	(0.339)	(0.019)
HP	2.821***	1.633***	-0.042***	0.002	-7.107***	0.466***
	(0.164)	(0.110)	(0.004)	(0.003)	(0.339)	(0.019)
random	2.527***	1.453***	-0.037***	0.004	-6.293***	0.465***
	(0.164)	(0.110)	(0.004)	(0.003)	(0.339)	(0.019)
Intercept (best)	93.934***	96.275***	1.533***	1.149***	42.517***	6.347***
	(0.116)	(0.078)	(0.003)	(0.002)	(0.240)	(0.014)
Observations	200	200	200	200	200	200
R^2	0.705	0.630	0.499	0.060	0.783	0.820
Adjusted R^2	0.701	0.625	0.492	0.045	0.779	0.818
Residual Std. Error	0.819(df = 196)	0.548(df = 196)	0.021(df = 196)	0.013(df = 196)	1.697(df = 196)	0.097(df = 196)
F Statistic	156.370*** (df = 3.0; 196.0)	111.378*** (df = 3.0; 196.0)	65.166*** (df = 3.0; 196.0)	4.161*** (df = 3.0; 196.0)	235.328*** (df = 3.0; 196.0)	298.438*** (df = 3.0; 196.0)
Note:	Group size: 10; maximum heuristic: 20; *p<0.1; **p<0.05; ***p<0.01					

Table 3

Regression analysis of the different types of groups with maximum heuristic of twelve on correlated landscapes. Each variable is a dummy variable for each type of group. The reference group type (intercept) is the group type of individually best-performing agents. The C coefficients are the coefficients for maximally C-diverse dummy variables similar to the HP coefficient, which is the coefficient for the maximally HP-diverse group type dummy variables. The columns represent the different performance indicators coupled with a specific mode of collaboration. Standard errors are reported in parentheses.

	outcome_corr, sequential	outcome_corr, non-sequential	checks_corr, sequential	checks_corr, non-sequential	moves_corr, sequential	moves_corr, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)
C	0.023	0.084	0.002	0.078***	1.361*	0.934***
	(0.172)	(0.177)	(0.004)	(0.005)	(0.789)	(0.052)
HP	0.021	0.132	0.003	0.081***	1.610**	0.980***
	(0.172)	(0.177)	(0.004)	(0.005)	(0.789)	(0.052)
random	-0.004	0.047	0.002	0.080***	1.567**	0.951***
	(0.172)	(0.177)	(0.004)	(0.005)	(0.789)	(0.052)
Intercept (best)	54.131***	55.089***	0.986***	0.685***	29.567***	3.155***
	(0.122)	(0.125)	(0.003)	(0.003)	(0.558)	(0.037)
Observations	200	200	200	200	200	200
R^2	0.000	0.003	0.003	0.703	0.028	0.721
Adjusted R^2	-0.015	-0.012	-0.012	0.698	0.013	0.717
Residual Std. Error	0.860(df = 196)	0.886(df = 196)	0.020(df = 196)	0.023(df = 196)	3.943(df = 196)	0.260(df = 196)

F Statistic	0.013 (df = 3.0; 196.0)	0.199 (df = 3.0; 196.0)	0.198 (df = 3.0; 196.0)	154.460*** (df = 3.0; 196.0)	1.877 (df = 3.0; 196.0)	169.241*** (df = 3.0; 196.0)
Note:	Assumptions for linear regression satisfied, Group size: 10; maximum heuristic: 12; *p<0.1; **p<0.05; ***p<0.01					

Table 4

Regression analysis of the different types of groups with maximum heuristic of twenty on correlated landscapes. Each variable is a dummy variable for each type of group. The reference group type (intercept) is the group type of individually best-performing agents. The C coefficients are the coefficients for maximally C-diverse dummy variables similar to the HP coefficient, which is the coefficient for the maximally HP-diverse group type dummy variables. The columns represent the different performance indicators coupled with a specific mode of collaboration. Standard errors are reported in parentheses.

	outcome_corr, sequential	outcome_corr, non-sequential	checks_corr, sequential	checks_corr, non-sequential	moves_corr, sequential	moves_corr, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)
C	0.191	0.418**	-0.011**	0.067***	-0.419	0.951***
	(0.174)	(0.176)	(0.004)	(0.005)	(0.726)	(0.059)
HP	0.126	0.400**	-0.010**	0.071***	-0.362	0.990***
	(0.174)	(0.176)	(0.004)	(0.005)	(0.726)	(0.059)
random	0.078	0.280	-0.009**	0.070***	0.059	0.970***
	(0.174)	(0.176)	(0.004)	(0.005)	(0.726)	(0.059)
Intercept (best)	56.019***	57.290***	0.976***	0.673***	28.329***	3.208***
	(0.123)	(0.124)	(0.003)	(0.004)	(0.513)	(0.042)

Observations	200	200	200	200	200	200
R^2	0.006	0.036	0.042	0.588	0.003	0.672
Adjusted R^2	-0.009	0.021	0.027	0.582	-0.012	0.667
Residual Std. Error	0.871(df = 196)	0.880(df = 196)	0.022(df = 196)	0.025(df = 196)	3.629(df = 196)	0.296(df = 196)
F Statistic	0.426 (df = 3.0; 196.0)	2.408^{*} (df = 3.0; 196.0)	2.838^{**} (df = 3.0; 196.0)	93.180^{***} (df = 3.0; 196.0)	0.228 (df = 3.0; 196.0)	134.090^{***} (df = 3.0; 196.0)
Note:	Assumptions for linear regression satisfied, Group size: 10; maximum heuristic: 20; $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$					

2a What is the association between group diversity and performance?

This research question targets broader patterns. Thus, it goes beyond the comparison of group types and compares overall associations of the two diversity measures with the three different performance indicators. This analysis is primarily done using visual analysis and correlation coefficients to quantify the linear relationship between diversity and performance. The former also underlines the appropriateness of correlation coefficients as it illustrates approximately linear relationships. The following subsection is structured by comparing the different performance indicators relative to each diversity measure considering all other dimensions.

C-diversity as well as the trial with the greater heuristic space score consistently more extreme on uncorrelated landscapes. Nevertheless, the results are very similar compared to their counterparts HP-diversity and trials with a space of twelve possible numbers for heuristics. The pattern is also clearly visible in Figures 1 (trial with heuristic space between one and twelve) and Figure 2 (trial with heuristic space between one and twenty). Yet, there is one exception: the sign flips for checks on the non-sequential mode of collaboration. While it is positive in Figure 1 and lower for C-diversity, it becomes negative with again more extreme extent for the case of C-diversity. These figures also show a fitted regression line for illustrative purposes, for further details see Appendix A for regression analysis. Interestingly, the baseline (the intercept) seems to be the driving difference with the coefficients often being less extreme for HP diversity. In conclusion, pattern indicates that C-diversity and the increased heuristic space lead to higher values for positive correlation coefficients and lower ones for negative values.

Besides, all correlations are positive regarding the outcome and rather mixed concerning efficiency. Diversity seems to be consistently positively associated with the outcome. However, the relative number of checks is only weakly negatively associated with diversity or even slightly positive in Figure 1 indicating the relevance of heuristic space. Additionally, the relative number of moves appears to be dependent upon the mode of collaboration: in the sequential mode diversity benefits efficiency while it is negatively associated in the tournament dynamics. Furthermore, regression analysis confirms the above-mentioned patterns (regression tables can be found in Appendix A). Hence, diversity improves the outcome whereas certain conditions need to be met for this to be the case concerning efficiency.

On correlated landscapes, the abovementioned pattern of amplified results for the trial with a greater heuristic space mostly prevails regarding the overall correlation between diversity and performance measurements. Concerning outcome, the results with the smaller heuristic space indicate opposing or inconclusive correlations compared to uncorrelated landscapes while diversity is slightly associated with higher outcomes with enlarged heuristic space. Similarly for the relative number of checks in the sequential mode it switches from diversity affecting efficiency negatively to contributing to it. Contrastingly, the correlation is consistently positive for the non-sequential mode and thus reflects the results on uncorrelated landscapes. The very same pattern concerning checks can be seen regarding the relative number of moves. Interestingly, the strong positive correlation is slightly weaker in the run with greater heuristic space and the pattern of switching the sign seen here with the sequential mode can be observed on the non-sequential mode on the uncorrelated landscape. These findings indicate complex interactions between the various factors at play.

Overall, there appear to be different dynamics on correlated landscapes with partially beneficial impacts of diversity particularly regarding outcome. A greater heuristic space coupled with diversity benefits outcome and partly efficiency. Diversity only contributes regarding the latter in the sequential mode of collaboration while it only becomes less damaging to efficiency if the space is greater for the non-sequential mode. These mixed and moderate findings for correlated landscapes as well as the partly decisive impact of enlarged heuristic spaces are supported by the regression analysis (see Appendix A).

Figure 1

The figure shows the association between diversity (x-axis) and performance indicator (y-axis) on uncorrelated landscapes for groups with agents whose maximum heuristic is twelve . Each datapoint represents one group and the colouring indicates which type of group the point belongs to. The processes of assembly/types of groups are maximally HP- and C-diverse (“HP”, “C”), individually best-performing (“best”) and randomly assembled (“random”). Data from 50 different landscapes is reported. Both diversity measures and modes of collaboration are shown. Furthermore, the regression line (blue) is fit for illustrative purposes. Also the correlation coefficient of the specific diversity measure and performance indicator are reported.

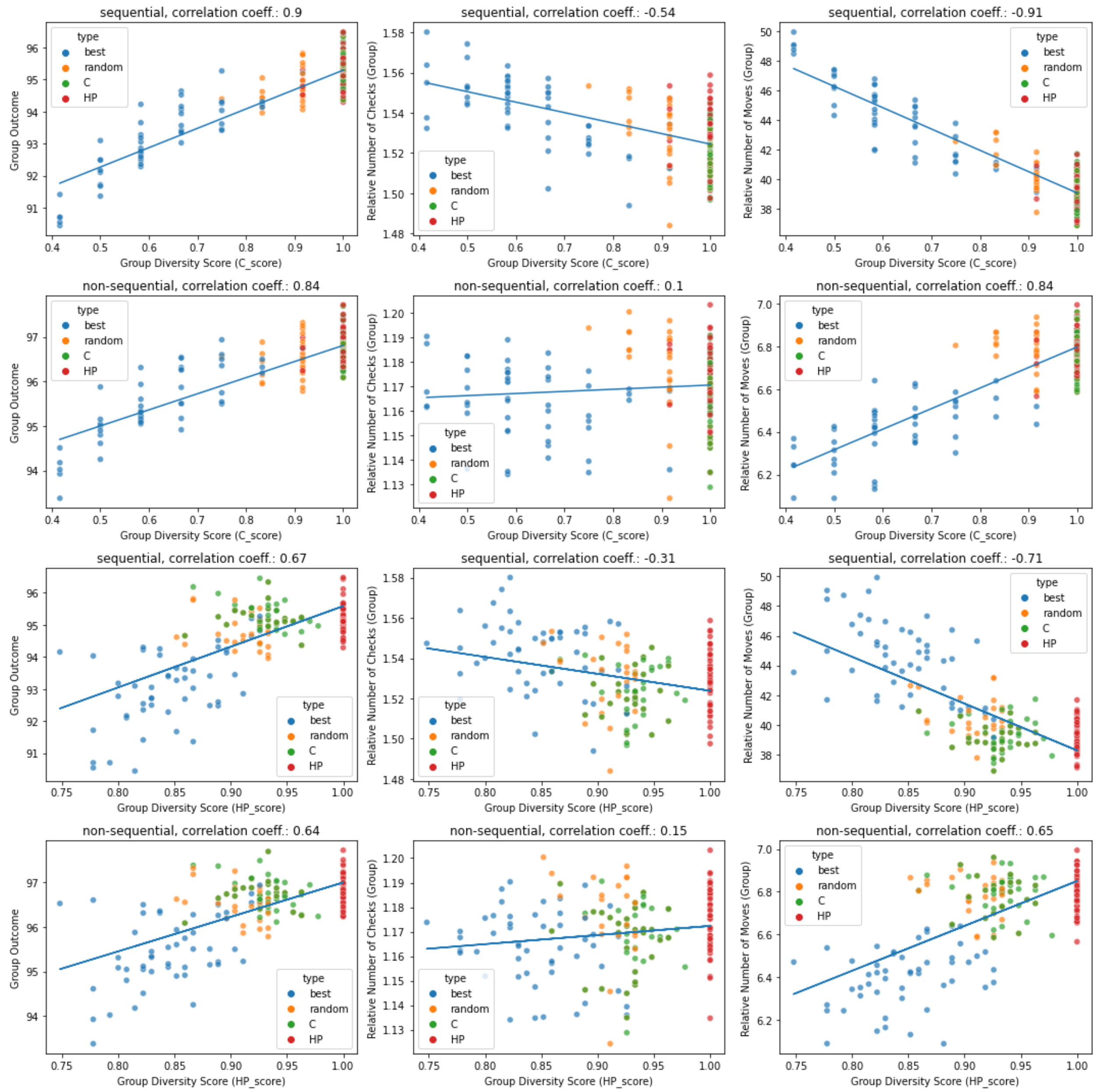


Figure 2

The figure shows the association between diversity (x-axis) and performance indicator (y-axis) on uncorrelated landscapes for groups with agents whose maximum heuristic is twenty. Each datapoint represents one group and the colouring indicates which type of group the point belongs to. The processes of assembly/types of groups are maximally HP- and C-diverse ("HP", "C"), individually best-performing ("best") and randomly assembled ("random"). Data from

50 different landscapes is reported. Both diversity measures and modes of collaboration are shown. Furthermore, the regression line (blue) is fit for illustrative purposes. Also the correlation coefficient of the specific diversity measure and performance indicator are reported.

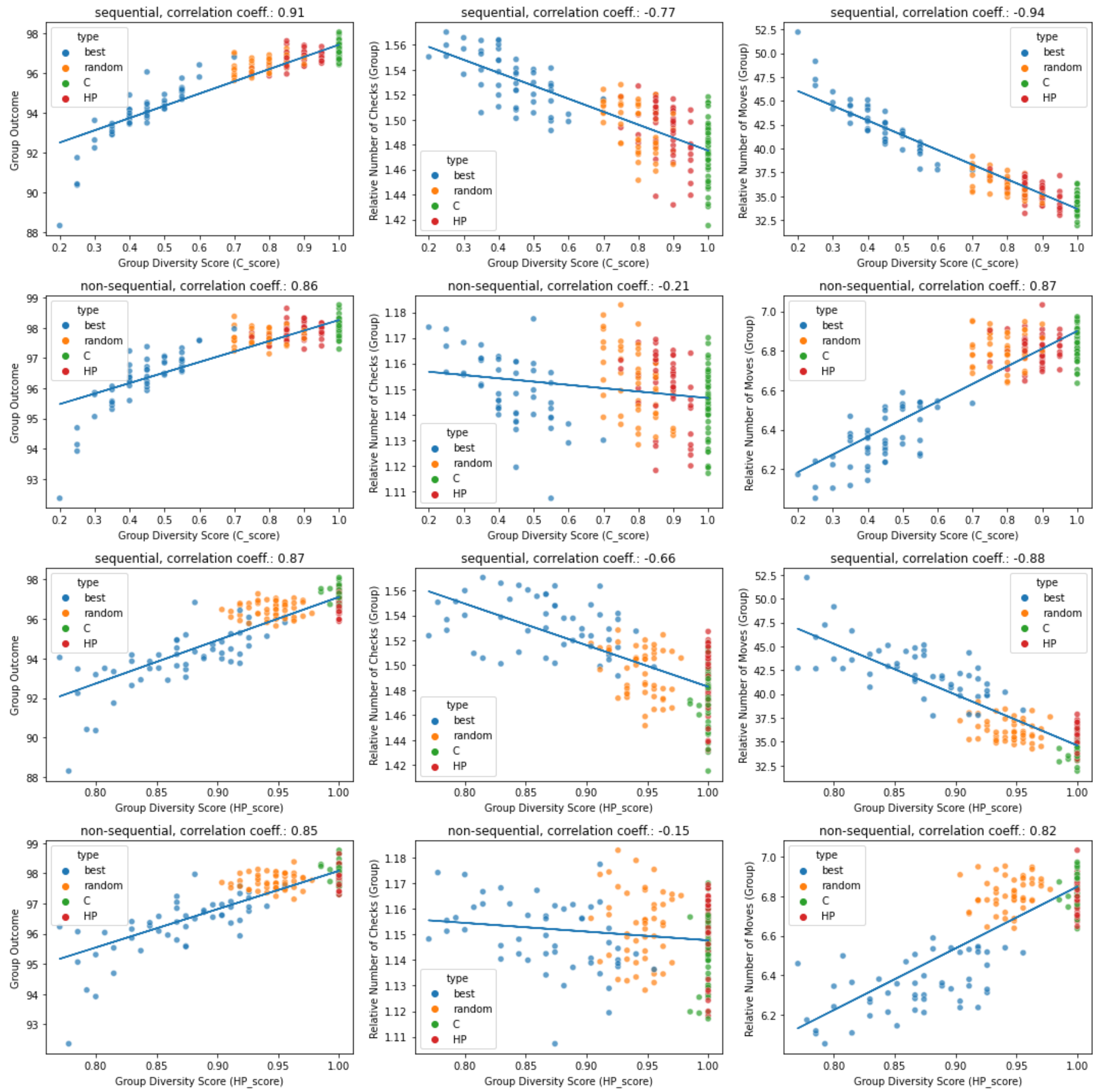


Figure 3

The figure shows the association between diversity (x-axis) and performance indicator (y-axis) on correlated landscapes for groups with agents whose maximum heuristic is twelve . Each *datapoint represents one group and the colouring indicates which type of group the point belongs to. The processes of assembly/types of groups are maximally HP- and C-diverse ("HP", "C"), individually best-performing ("best_corr") and randomly assembled ("random"). Data from 50 different landscapes is reported. Both diversity measures and modes of collaboration are shown. Furthermore, the regression line (blue) is fit for illustrative purposes. Also the correlation coefficient of the specific diversity measure and performance indicator are reported.*

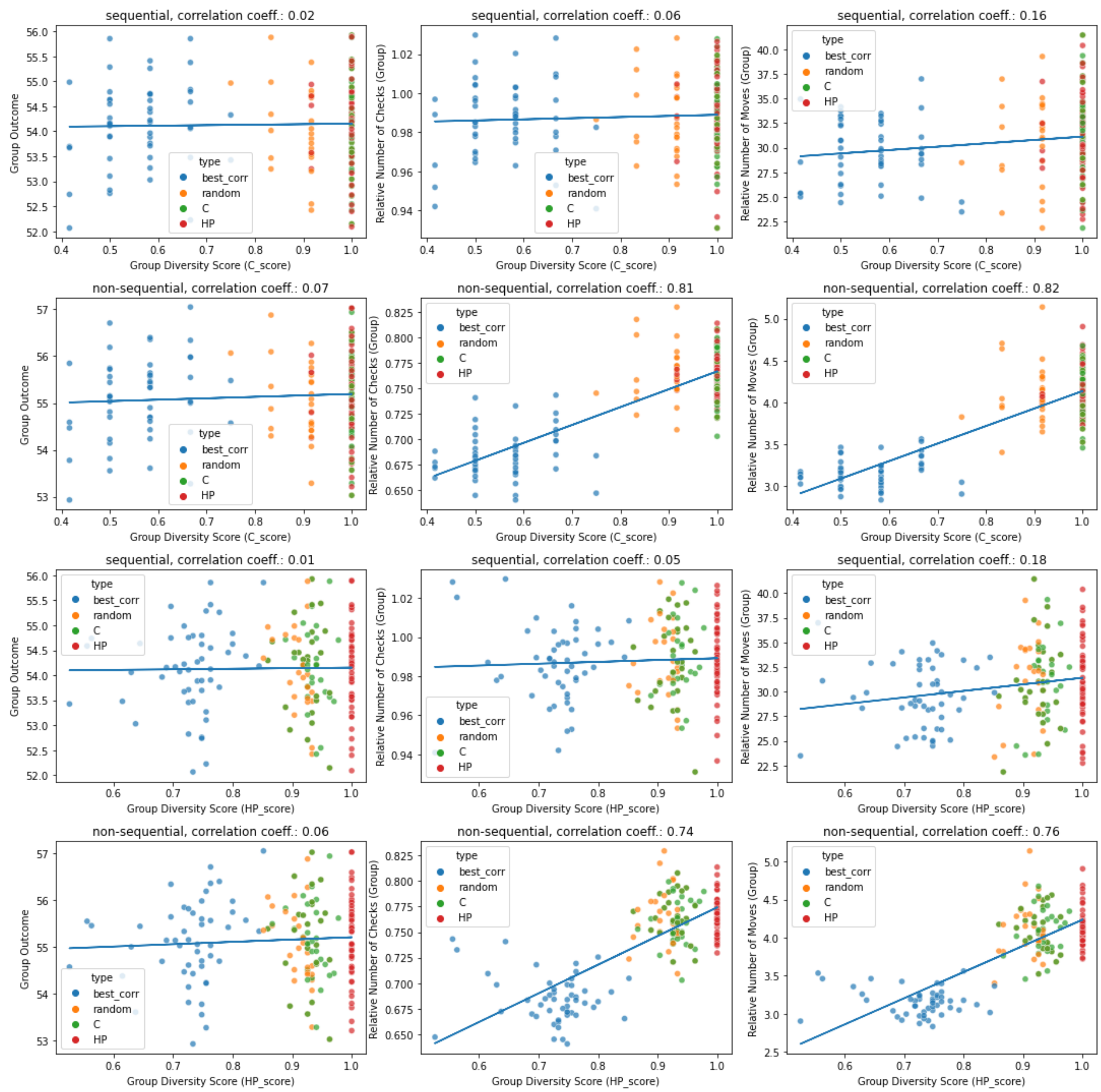
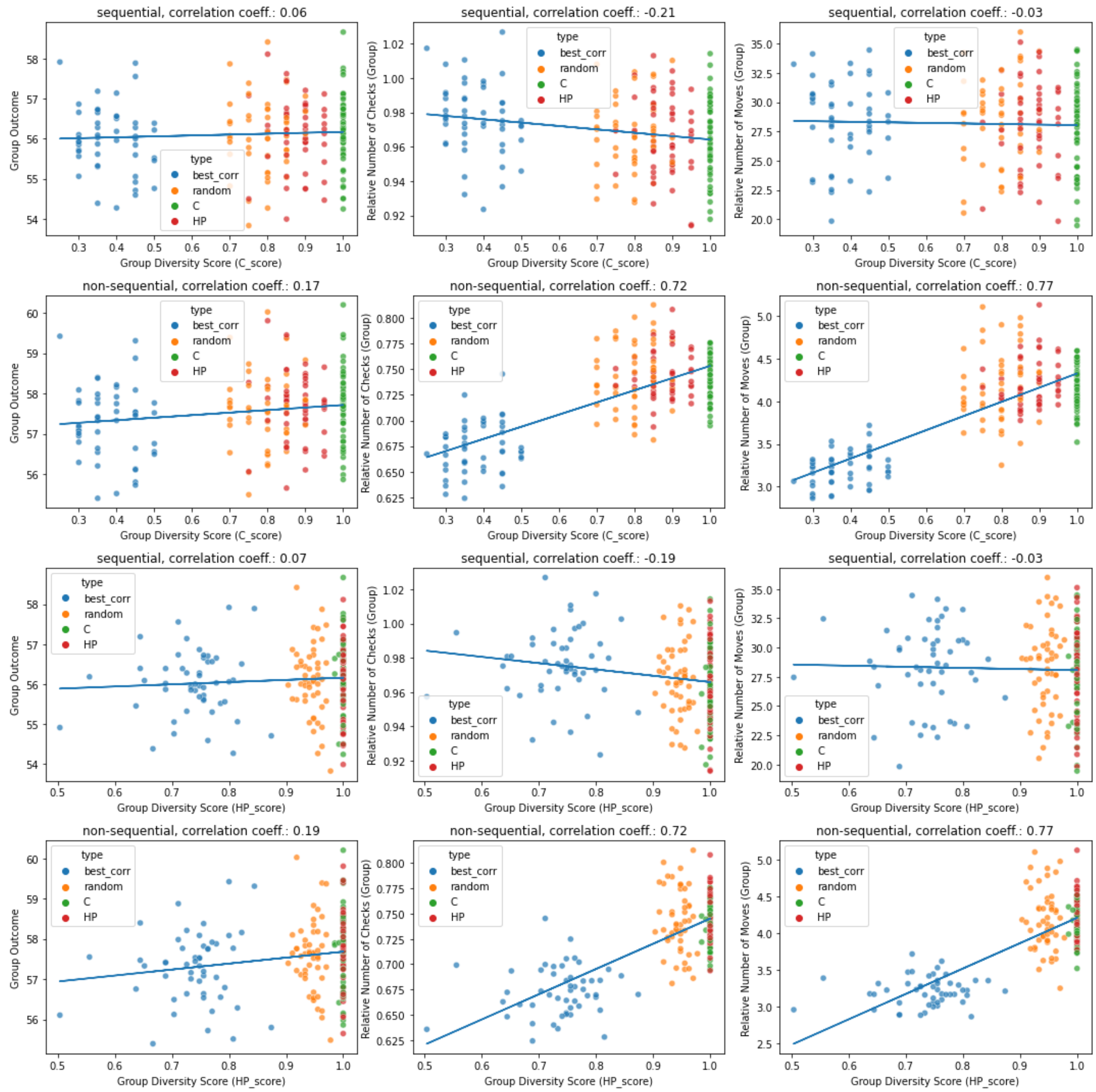


Figure 4

The figure shows the association between diversity (x-axis) and performance indicator (y-axis) on correlated landscapes for groups with agents whose maximum heuristic is twenty. Each datapoint represents one group and the colouring indicates which type of group the point belongs to. The processes of assembly/types of groups are maximally HP- and C-diverse ("HP", "C"), individually best-performing ("best_corr") and randomly assembled ("random"). Data from 50 different landscapes is reported. Both diversity measures and modes of collaboration are shown. Furthermore, the regression line (blue) is fit for illustrative purposes. Also the correlation coefficient of the specific diversity measure and performance indicator are reported.



2b What is the association between average individual ability and performance?

This subsection explores the association between average individual ability (outcome) and performance indicators. This aspect has not been explored explicitly. Here the analysis is done for each of the performance indicators as well as other domains. The results paint a multifaceted and complex picture of the impact of average individual ability.

The outcome is influenced very distinctly by the average individual ability depending on the type of landscape. On uncorrelated landscapes, the association is negative. However, it appears to be neutral or even positive for all the groups, but the individually best-performing group type, which has a lower average and much wider spread dragging the overall association down. This effect is consistent but stronger in the sequential mode and the trial with larger heuristic space. On correlated landscapes, the association is strongly positive with two parallel distributions: one with the individually best-performing and the other with the remaining group types, but both with strongly positive relationships, which are weaker for the trial with larger heuristics space and amplified in the sequential mode. Hence, depending on the landscape, high individual ability can be an asset or a liability.

The pattern of influence of landscape is repeated with the mode of collaboration for efficiency. The relative number of checks and moves increases with increased average individual ability in the sequential mode of collaboration whereas it decreases or stays steady in the tournament-mode. Here, the results are amplified on correlated landscapes. Interestingly, the driver of the results often is the difference between individually best-performing group type and the others, as they tend to cluster either below or above giving it the decisive direction.[8] Thus, the differences between these types of groups are the significant factor mostly, determining the association.

Figure 5

The figure shows the association between average individual outcome in a given group(x-axis) and performance indicator (y-axis) on uncorrelated landscapes for groups with agents whose maximum heuristic is twelve . Each datapoint represents one group and the colouring indicates

[8] Exceptions are non-sequential+uncorrelated for checks and sequential+correlated for both

which type of group the point belongs to. The processes of assembly are maximally HP- and C-diverse ("HP", "C"), individually best-performing ("best") and randomly assembled ("random"). Data from 50 different landscapes is reported. Both modes of collaboration are shown. Furthermore, the regression line (blue) is fit and added for illustrative purposes. Also, the correlation coefficient of the average individual outcome and performance indicator are reported.

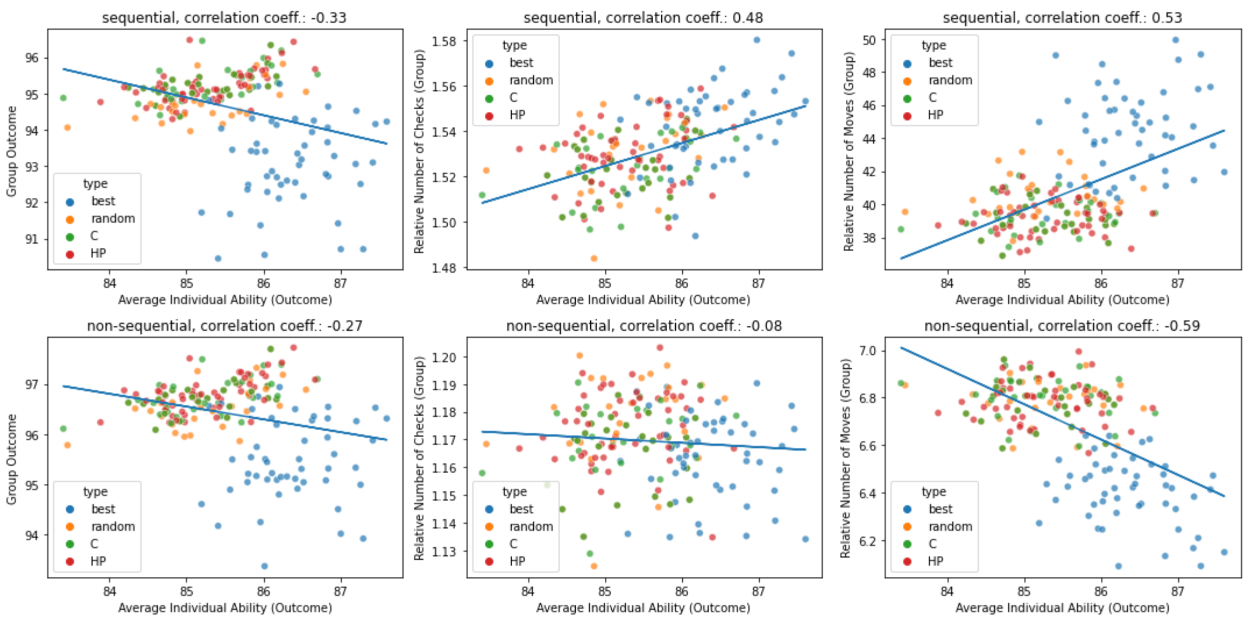


Figure 6

The figure shows the association between diversity (x-axis) and performance indicator (y-axis) on uncorrelated landscapes for groups with agents whose maximum heuristic is 20. Each datapoint represents one group and the colouring indicates which type of group the point belongs to. The processes of assembly are maximally HP- and C-diverse ("HP", "C"), individually best-performing ("best") and randomly assembled ("random"). Data from 50 different landscapes is reported. Both modes of collaboration are shown. Furthermore, the *regression line (blue) is fit and added for illustrative purposes. Also, the correlation coefficient of the average individual outcome and performance indicator are reported.*

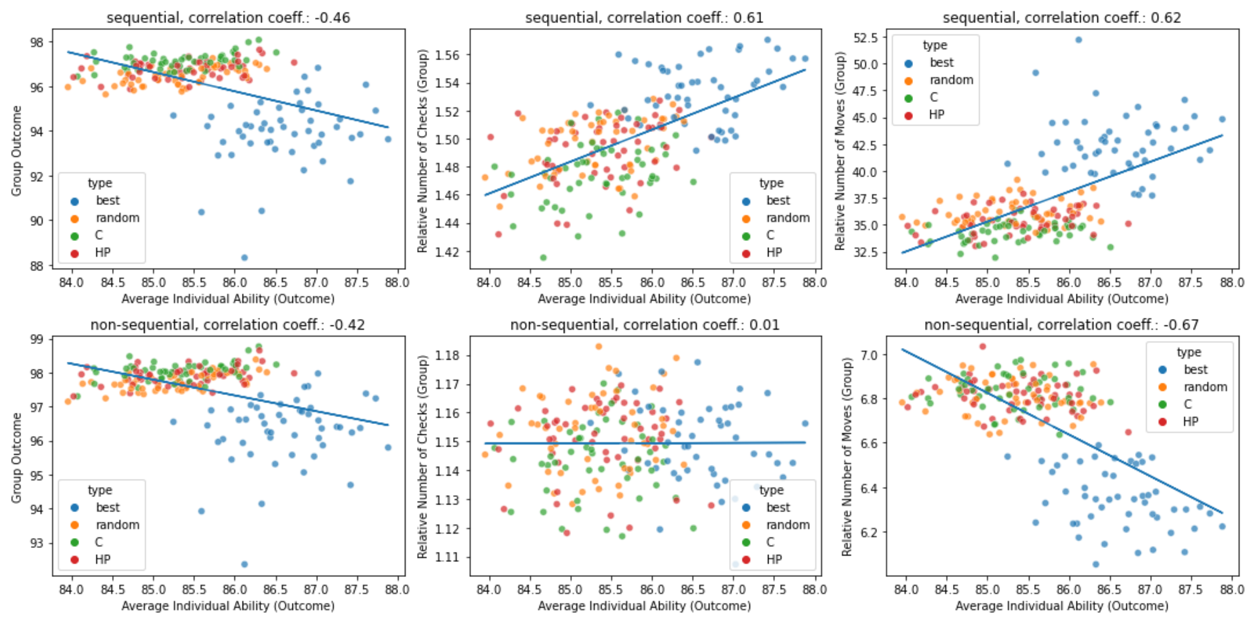


Figure 7

The figure shows the association between diversity (x-axis) and performance indicator (y-axis) on correlated landscapes for groups with agents whose maximum heuristic is twelve. Each datapoint represents one group and the colouring indicates which type of group the point belongs to. The processes of assembly are maximally HP- and C-diverse ("HP", "C"), individually best-performing ("best_corr") and randomly assembled ("random"). Data from 50 different landscapes is reported. Both modes of collaboration are shown. Furthermore, the regression line (blue) is fit and added for illustrative purposes. Also, the correlation coefficient of the average individual outcome and performance indicator are reported.

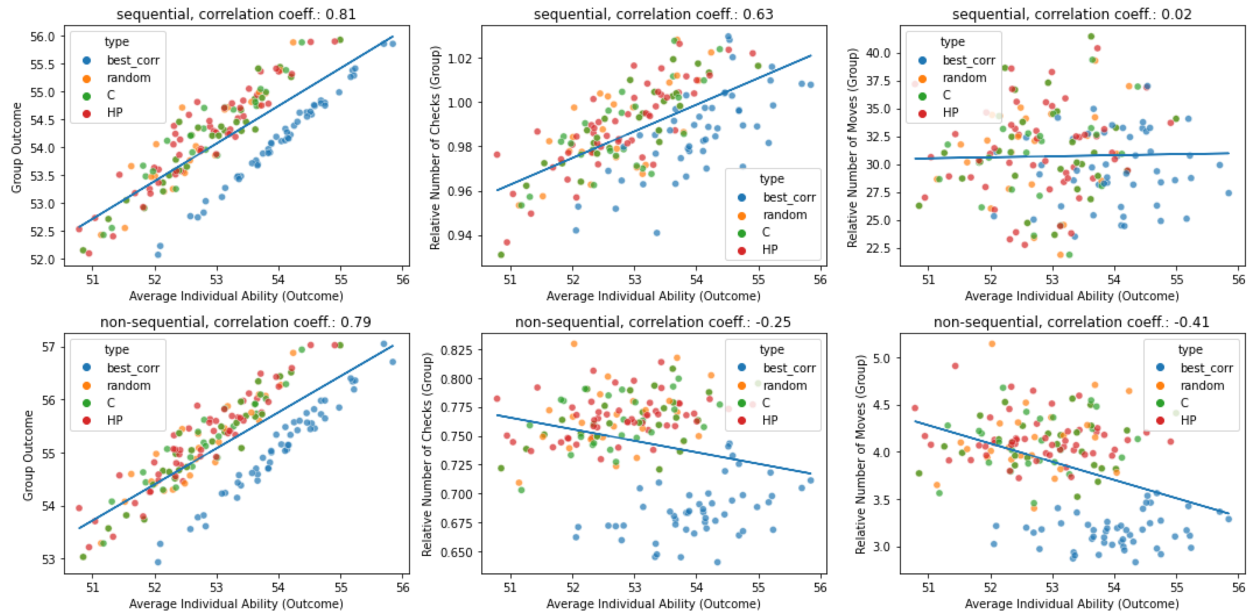
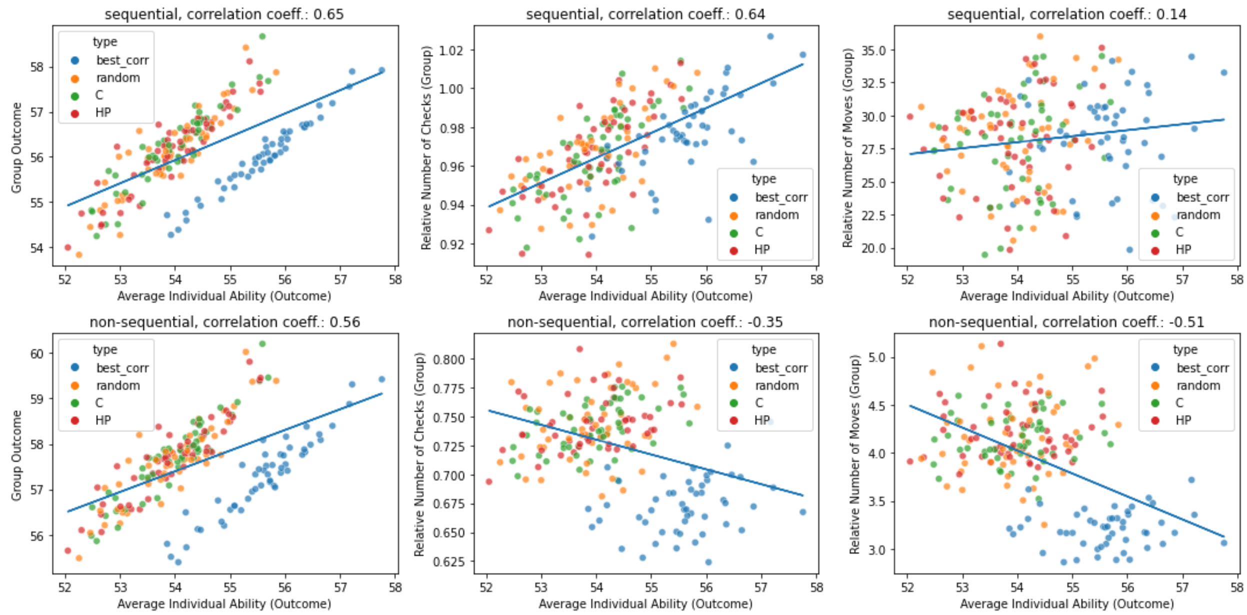


Figure 8

The figure shows the association between diversity (x-axis) and performance indicator (y-axis) on correlated landscapes for groups with agents whose maximum heuristic is 20. Each datapoint represents one group and the colouring indicates which type of group the point belongs to. The processes of assembly are maximally HP- and C-diverse ("HP", "C"), individually best-performing ("best_corr") and randomly assembled ("random"). Data from 50 different landscapes is reported. Both modes of collaboration are shown. Furthermore, the regression line (blue) is fit and added for illustrative purposes. Also, the correlation coefficient of the average individual outcome and performance indicator are reported.



2c What is the association between the group properties as well as their interactions and performance?

This research question investigates the interplay between performance and both group properties. Hence, this subsection is once again structured according to the three different performance indicators and the different domains to describe how group diversity and average individual ability are associated with group performance.

The outcome is heavily affected by an interplay between both group properties with varying importance of each depending on the landscape type. On uncorrelated landscapes the group diversity is highly associated with group outcome: groups with low diversity have not outperformed the ones with higher irrespective of individual ability. However, higher individual ability is associated with improvements of the outcome among the highly diverse groups. The latter statement remains true on correlated landscapes, but the former does not. That means that individual ability becomes the more important predictor of group outcome. Nevertheless, a lower average individual ability coupled with high diversity can still reach the same, if not higher group outcome. These findings hold across deliberation mode and diversity measure although they appear to be slightly clearer for the lower heuristic space.

Efficiency does not show patterns as clear and also more complex impacts of the different dimensions. On correlated landscapes, higher diversity and individual ability together are positively associated with group outcome although the influence of ability in the non-sequential mode is limited and in the sequential mode there are some groups which are highly efficient while being very uniform. Thus, ability seems to be a better predictor in the sequential mode. Interestingly, the efficiency often peaks at a diversity score of about 0.8-0.9, so not at the extreme end of the spectrum. On uncorrelated landscapes, the results are highly dependent on the deliberation mode. While the combination of high individual ability and low diversity is a winning formula for the sequential mode, the opposite is true for the tournament mode. For the latter mode, extreme diversity is not necessary with moderate diversity scores coupled with low individual ability being sufficient very often. Interestingly, the dynamics for the number of checks is rather inconclusive here although the results are clear regarding the number of moves. Thus, ability and diversity together create their own dynamics but these are very much dependent upon the type of landscapes and partially also on the mode of deliberation.

Figure 9

The figure shows the association between diversity (x-axis), average individual outcome (y-axis) and performance indicator (colouring) on uncorrelated landscapes for groups with agents whose maximum heuristic is twelve. Each hexagon averages the datapoints' (groups') performance falling into that area. Data from 50 different landscapes is reported. Both modes of collaboration and diversity measures are shown.

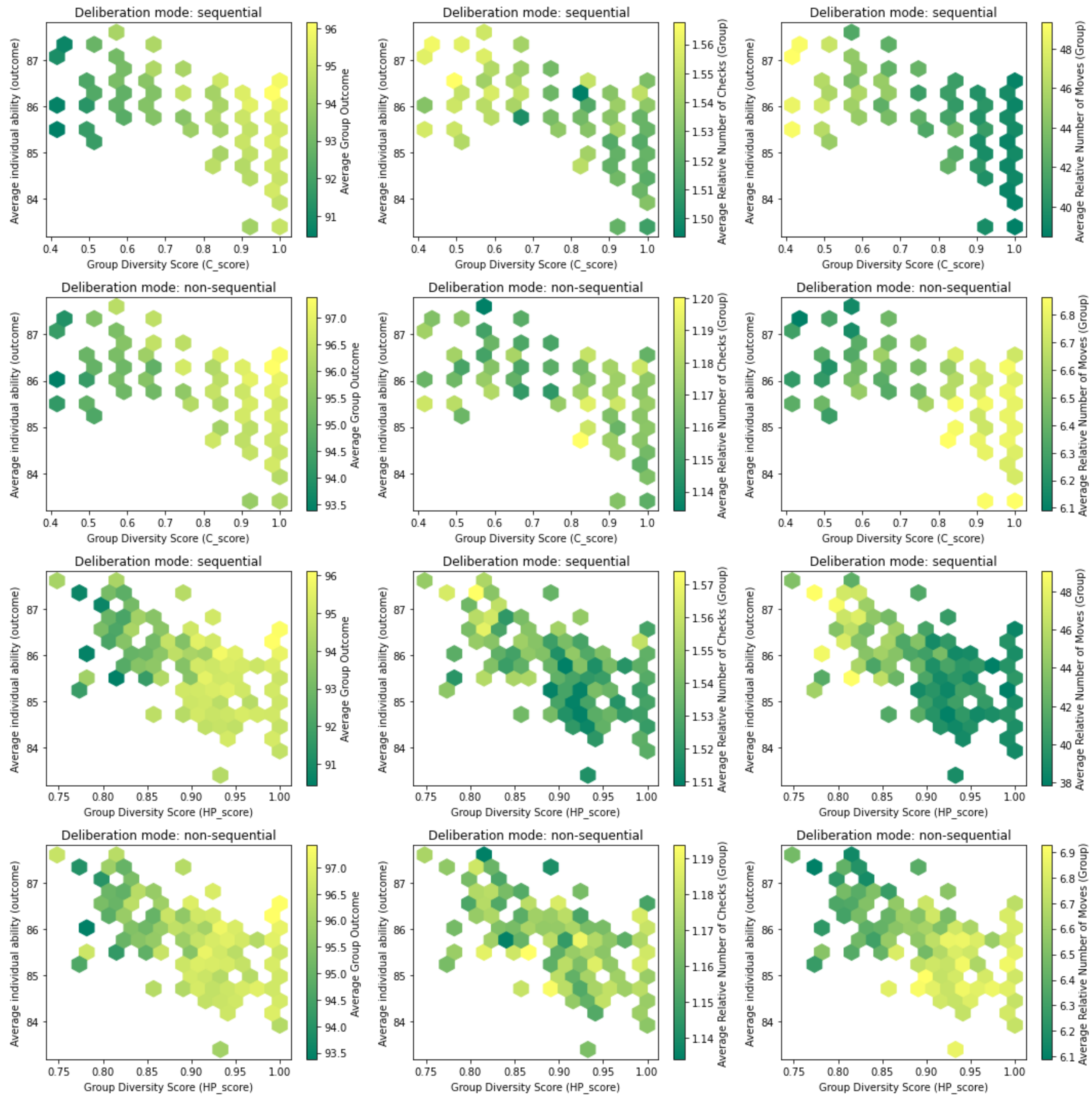


Figure 10

The figure shows the association between diversity (x-axis), average individual outcome (y-axis) and performance indicator (colouring) on correlated landscapes for groups with agents whose maximum heuristic is twelve. Each hexagon averages the datapoints' (groups') performance falling into that area. Data from 50 different landscapes is reported. Both modes of collaboration and diversity measures are shown.

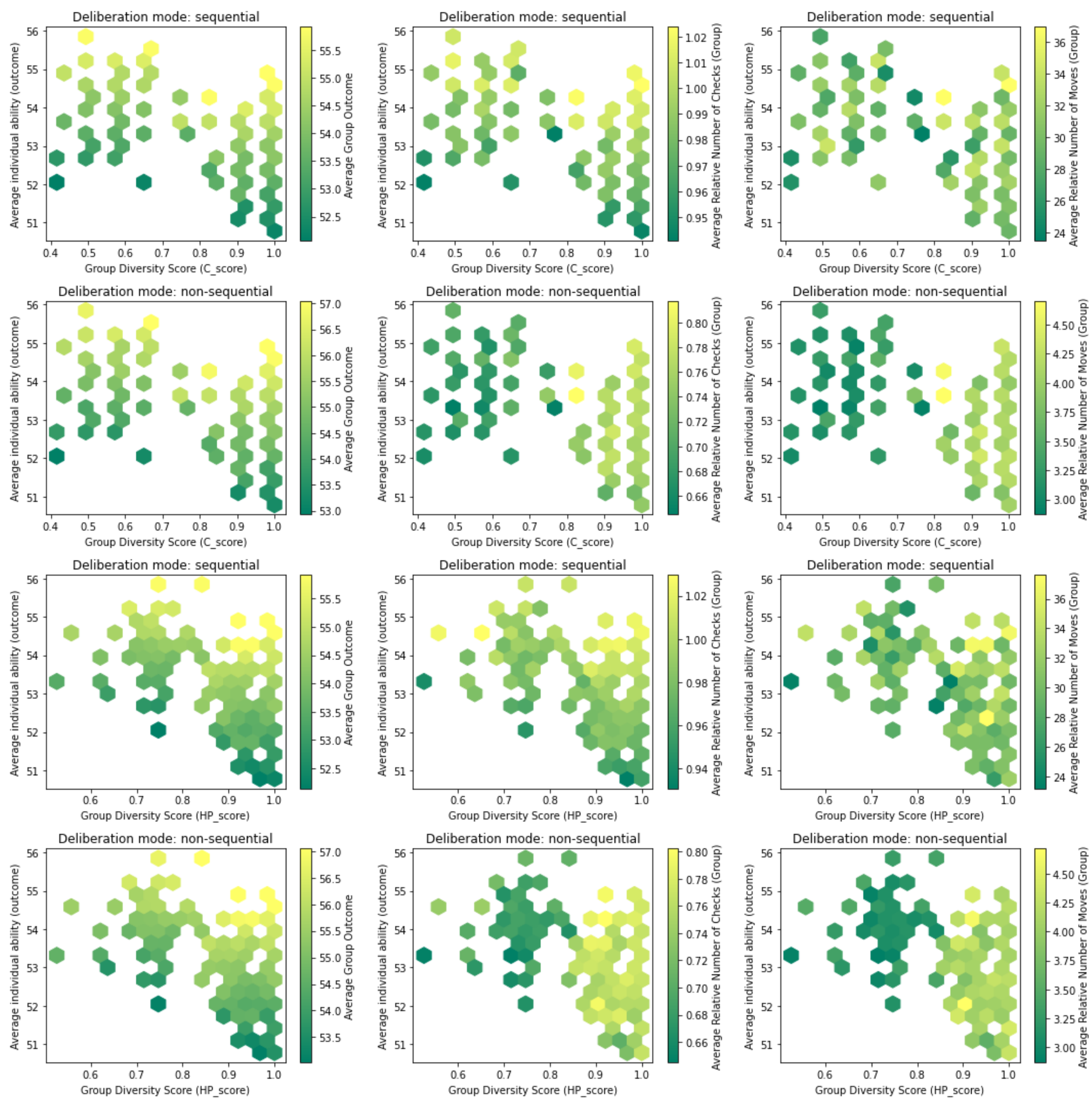


Figure 11

The figure shows the association between diversity (x-axis), average individual outcome (y-axis) and performance indicator (colouring) on uncorrelated landscapes for groups with agents whose maximum heuristic is 20. Each hexagon averages the datapoints (groups) performance falling into that area. Data from 50 different landscapes is reported. Both modes of collaboration and diversity measures are shown.

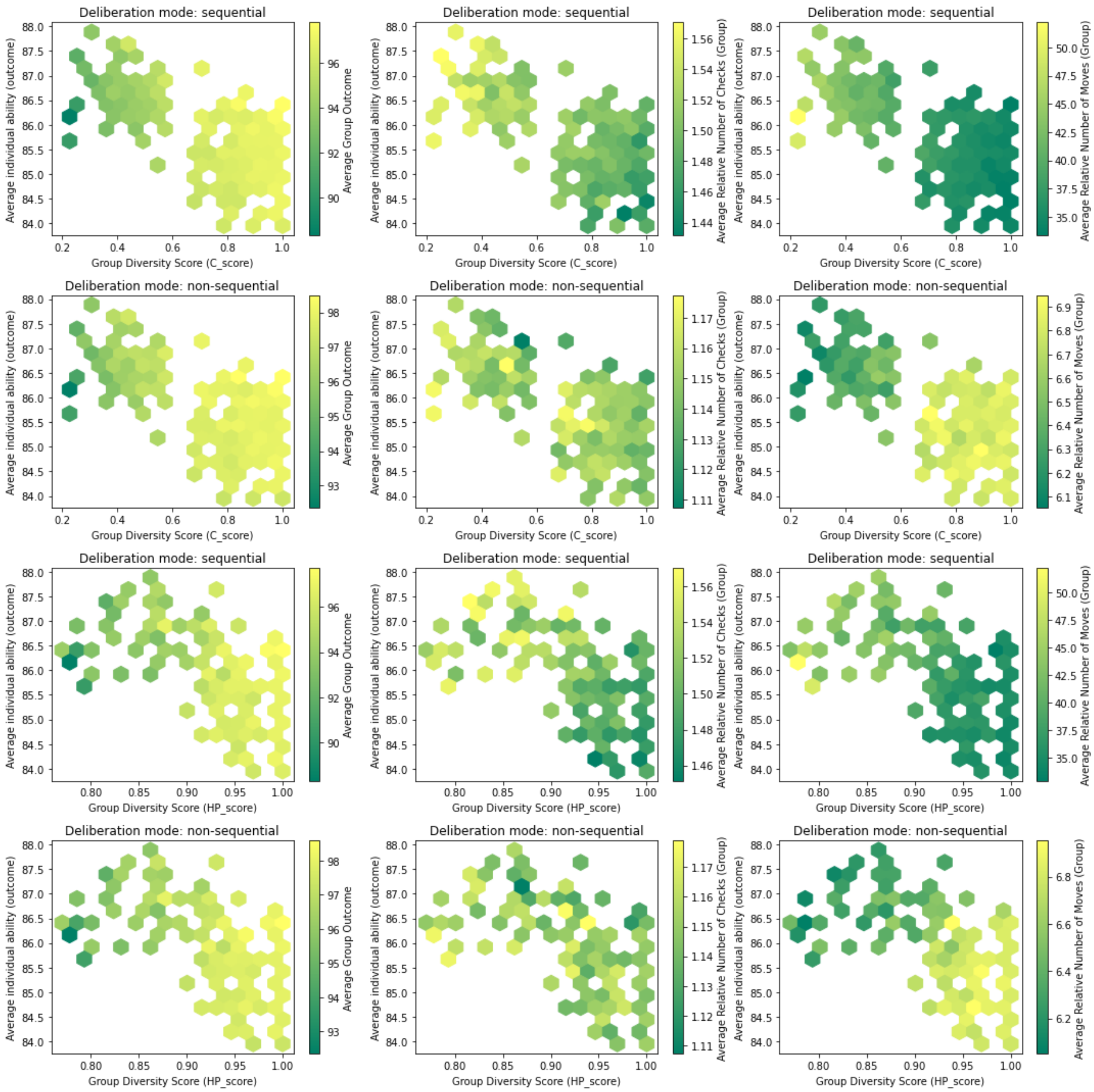
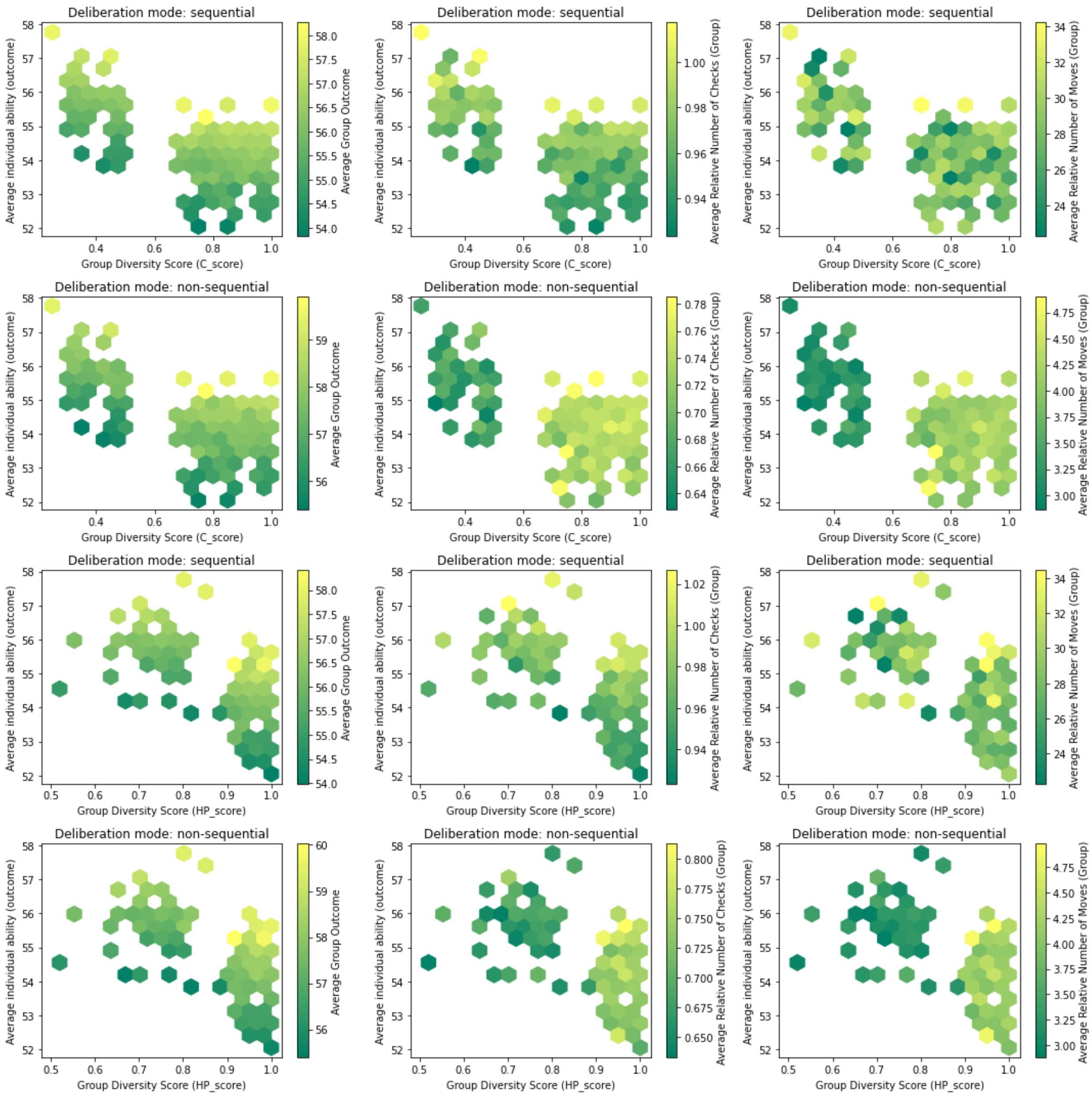


Figure 12

The figure shows the association between diversity (x-axis), average individual outcome (y-axis) and performance indicator (colouring) on correlated landscapes for groups with agents whose maximum heuristic is 20. Each hexagon averages the datapoints' (groups') performance falling into that area. Data from 50 different landscapes is reported. Both modes of collaboration and diversity measures are shown.



Discussion

This section aims at providing an overview of the results and interpretation thereof by connecting them to relevant literature as well as broadening the view on them. Firstly, the results concerning the final outcome are reviewed, interpreted and compared to earlier findings mentioned in the literature review. Secondly, the novel aspects of efficiency are connected and explained as far as possible. Thirdly, the broader implications of modelling assumptions are discussed and contrasted with other modelling approaches from domain of epistemic value of diversity. Hence, this section combines several views derived from various sources to provide a wide-ranging basis for conclusions.

Interpreting Results & Potential Explanations

Generally, the results of previous research can be confirmed on the basis of the results presented above. These include that C-diversity as a group property is a better predictor of outcome than HP-diversity (Singer, 2019). This is also reflected in the finding that maximally C-diverse groups outperform any other group on uncorrelated landscapes, which is a novel yet to be expected insight. Interestingly, it also appears to be a better predictor for efficiency-metrics. Further, the diminishing value of diversity vs expertise on correlated landscapes (Holman et al., 2018; Grim et al., 2018) is confirmed. Also, the impact of the heuristic space on outcome found in previous research (Holman et al., 2018; Grim et al., 2018) could be confirmed. This means that when considering a notion of diversity, it is more relevant to think of it as maximising the range of relevant skills to maximise the outcome.

This stands in some contrast to the finding that transferable expertise, meeting the minimal criterion, appears to be more relevant with increasingly predictable problems (Holman et al., 2018; Grim et al., 2018), which I have found as well. However, the results show that regardless of the level of diversity, expertise is associated with an increase in the final outcome. Hence, this finding counters the claim that diversity trumps ability as diversity coupled with ability is the winning formula with greater importance of one or the other depending on the type of task and to a lesser extent the mode of collaboration, and heuristic space. These findings might explain why others have found that a mix of experts and non-experts in one group is often the superior composition (Grim et al., 2018).

The consistent relevance of individual expertise weakens the independence thesis as properties of an epistemic group (here diversity and average individual ability) can differ from the properties of individual members (Mayo-Wilson et al., 2011). This claim does hold true for diversity but is always impacted by the members through their individual ability. Nevertheless, considering the importance of heuristic space or in other words, the range of relevant skills, diversity can hold great value in a variety of unpredictable settings requiring a wide range of skills as it counters the relative value of expertise on correlated landscapes. Additionally, the group type of the individually best-performing agents are much less diverse, which in turn is associated with lower group outcome due to the underperformance of this group (Hong & Page, 2004). Hence, diversity in some cases should be prioritised over individual ability when one seeks to optimise outcome.

Also, the impact of the deliberation dynamics on the outcome found previously can be confirmed. While Hong and Page (2004) state that they found overall comparable results, others found differences of when DTA on correlated landscapes as well as the general magnitude of findings (coefficients) with advantages of diversity in the tournament-mode (Grim et al., 2018). Thus, the superiority of diversity or transferable expertise depends to some extent on the mode of collaboration.

However, the results for the efficiency metrics completely depend upon the deliberation dynamics. This finding is not very surprising since their key differences lie within how they collaborate which is likely to have some impact on the outcome but much more on the efficiency since both are procedural aspects. In concrete terms, the number of checks in the non-sequential mode must be much higher in absolute numbers since each agent arrives at their local optimum to “share” the result with the group and pick the best result. This is indeed what can be observed in the absolute number of moves and checks (see appendix B for an example). As a likely consequence the number of moves is also higher since they are more likely to keep making advances, which can also be seen based on the better outcome. These speculations are reflected in the findings.

Contrastingly, the results reverse themselves when *relative* efficiency is taken into account. There the baselines for the non-sequential modes are lower regarding both efficiency measures while the slopes are negative for the sequential mode. These results show also that the group type with individually best-performing agents is overall less efficient than the more diverse group types regarding the sequential mode of collaboration. Yet, in the tournament mode, where every group needs more moves and checks, this is not the case anymore as the groups need less moves on average, which affects the relative standing of the diverse group types. Additionally, the outcomes are also higher with smaller gaps between the group type of best-performers and the diverse ones in the non-sequential mode, which leads to less relative efficiency differences if they roughly need the same number of moves and checks. These contradictory findings are highly relevant as they show the importance of the absolute number of actions which need to be taken versus the relative.

Arguably, the resources needed to achieve a given outcome, in this case reaching a local optimum, is often more likely to be relevant as resources are generally scarce. In other words, relative efficiency constitutes a more relevant object of inquiry as it is more relevant in many practical settings as the debate on satisficing shows (Katsikopoulos, 2014). Regarding this relative efficiency, I can show that deliberation dynamics affect the conclusion which can be drawn. Diversity and ability do affect efficiency. This becomes apparent when groups are compared to the group type of individually best-performing agents. Hence, whether a given group property is beneficial depends less on the inherent efficiency of a given group but rather on the efficiency of the reference group type (individually best-performing), which is dependent upon the form of deliberation.

Comparing ability with diversity on relative efficiency, the opposite picture reveals itself. In the non-sequential mode, the association is neutral or negative while it is positive in the tournament-mode. Thus, depending on the mode, ability is beneficial for efficiency (tournament) or not (sequential). The differences between ability and diversity may occur due to the opposing nature of both group properties with high performers often being alike (less diverse). This would also explain why the results are often driven by stark contrasts between the types of groups with the cluster of individually best-performing groups being apart from the other groups regarding diversity and individual ability, which in turn determines the nature of the association. This is also seen in the combined analysis where one of the two group properties is high with the other one being minimal among the highly efficient groups on uncorrelated landscapes (although a mix is once again best for correlated ones). Overall, one needs to weigh the benefits of diversity and ability very differently depending on the mode of collaboration but also the type of landscape.

Modelling Assumption

Before reaching a conclusion, it is worth questioning the model. The model itself offers plenty of opportunity to model different tasks, a variety of agents and deliberation dynamics. Simultaneously, all these factors remain highly abstract and simplified. Yet, this aspect is not necessarily problematic but should rather be considered an opportunity to gain a deep understanding of the variables at play due to the simplicity making that possible. Grasping all variables and their impacts and interactions already constitute a significant challenge. Thus, the criticism of certain aspects expressed in the following does not serve the purpose of undermining the relevance of the findings. Instead, the paragraphs below should raise awareness considering potential conclusions and further use of the findings in other fields.

The main criticism one may voice could be the deterministic nature of the model. This aspect is apparent on many different levels. When agents deliberate there is no cost of communication or communication considered at all, no learning (Hong & Page, 2004), no loss of information and no disagreement. Hence, one could describe the model as highly artificial. One would not expect full truthfulness, especially in groups where experts are present. They might value their own proposal potentially higher and push them accordingly. Similarly, each agent enjoys full credibility from each other agent. Additionally, they have the advantage of having full access to all findings from the other group members without any loss of information or penalty imposed on bigger or more diverse groups. The latter aspect should be seen as especially concerning since it could be shown that culturally diverse groups, which are also more functionally diverse (Nisbett & Ross, 1980; Thomas & Ely, 1996), need to cope with initial communication issues (Macleod, 1996; Polzer et al., 2002; Watson et al., 1993). This may also be the case with “only” functionally diverse groups with contributions from a variety of fields which may not be known to everyone in a respective group. All these assumptions are rather unrealistic to occur in genuine group deliberations.

Furthermore, not only the group dynamics to solve a task can be questioned but also the task itself. The modelling of the task has already evolved from completely random (Hong & Page, 2004) to correlation of adjacent positions (Grim et al., 2018; Holman, 2018). While it fulfils the criterion of being a challenging task i.e. not consistently solvable by any agent (Hong and Page, 2004), it may be artificial to keep looping over it and having only one smoothing factor/level of correlation as tasks at different stages in bigger projects potentially require varying expertise. For instance, at an early stage, randomness simulating unpredictability and the need for a wide variety of skills may later be substituted for a greater need of specific transferable expertise. Thus, the tasks posed for agents may be too monotonous and inflexible. While variability of tasks could be simulated in different trials implying a change of the group it may still be relevant for cases where such a switch of group members may not be possible.

Finally, agents themselves may be misrepresented if they were to resemble humans and their skill sets. First, the number of skills may not be sufficiently represented by three heuristics depending on how wide the notion of heuristic is understood. However, potentially more profoundly, the lack of variability appears to be even less realistic as very proficient problem solvers should have a greater variety of relevant skills, perhaps particularly in one area if they are experts in a specific field. Thus, the representation of agents could be criticised.

Nevertheless, careful considerations need to be made to ultimately judge on the value of these potential additional features. A simulation from which very practical insights can be derived needs to be structured in very specific ways. To achieve such specificity the target system i.e. the context to which it will be applied, needs to be studied (Alexandrova & Northcott, 2009). Since the paper by Hong and Page (2004) does not make such specific references and aims at a rather general argument the lack of detailed implementation and targeted design can be considered an advantage and starting point for further research as the example for representative democracies shows (Grim et al., 2020).

Beyond Hong & Page

The model of epistemic communities originally developed and later advanced and presented here is not the only one of its kind. There are other examples of such models, which have some of the aspects touched upon above incorporated (Alexander et al., 2015; Sakai, 2020;

Thoma, 2015; Weisberg & Muldoon, 2009; Zollman, 2007, 2010). Thus, the following subsection describes one such alternative briefly to discuss how the model primarily used throughout this thesis can be adapted, specified and combined. These alternative approaches show the importance and opportunities for computational modelling in general and these approaches in particular.

The examples considered here are the models developed by Zollman (2007, 2010). These models target scientific communities as a prime example of epistemic group dynamics. Thus, specific assumptions about the underlying mechanisms can be made and reasoned for. An instance of this explicitness is the basis for individual decisions: the next move is made based on the expected payoff from previous results, a reasonable assumption in academic research (Zollman, 2007). Additionally, this also resolves the issue of reporting truthfully as the punishment would be severe, yet there have been documented cases. Hence, the study of the target system makes the results less generalisable yet more precise and increases credibility.

Furthermore, two relevant aspects are incorporated into the models. Firstly, the ever-present issue of uncertainty and subjective experience is captured by using a Bayesian modelling approach representing findings and viewpoints thereof (Zollman, 2007, 2010). This eliminates the determinism present in the model proposed by Hong and Page (2004) and adds opportunities to represent different viewpoints through varying priors. This aspect is intrinsically linked to the access to evidence. Findings from other members of the epistemic community reach a given agent only through their neighbours. Thus, adjusting the network structure can yield different results and provide insights into the importance of connectivity (Zollman, 2007, 2010). Using Bayesian modelling in different network scenarios, it is found that higher connectivity leads to quicker conclusion but also to a higher degree of false conclusion, which can be combated by extreme priors i.e. steadfast members (Zollman, 2010).

These findings suggest that diversity of views plays a significant role in the findings provided by Zollman (2007, 2010), similar to the diversity of problem-solving heuristics and perspectives (Hong & Page, 2001, 2004). The diversity is achieved through scientists who either pursue their next moves based on limited information and/or extreme ‘ideas’. The former indicates that subjective views are steadier due to a potential lack of conflicting evidence while the latter creates more resistant agents potentially providing evidence to steer the group overall into the epistemically right direction.

These approaches yield potential for improvements of the model primarily discussed here. The limited access to information, agents with beliefs generally uncertainty constitute highly relevant aspects. Thus, implementing such mechanisms into the model originally proposed by Hong and Page (2004) at each individual move could be an interesting option to pursue and increase relevance of the findings based on the specific interests and target system at the cost of simplicity.

With this alternative model and the brief introduction of a potential merge, I aim to show that alternatives are possible, have justifications on their own, and are not necessarily contradicting yet also not necessarily beneficial. Both models find advantages of diversity in different ways. Additionally, their setup makes differing aspects apparent and hence links findings to these. This creates open questions for potential combinations and whether both models were to come to the same conclusions present in the other. Contrastingly, combining both may blur the findings as the results may not be attributable to one feature or another.

Limitations

There are several limitations to this study beyond the ones inherent in ABM. Some of these limitations are outlined above, particularly the ones which are present due to the general setup. Simultaneously, potential remedies and advantages of these shortcomings are addressed. Overall, these limitations can all be tackled in case it is deemed necessary e.g. after studying the target system on which insights should be generated (Alexandrova & Northcott, 2009). This leads also to the result that the findings provided here are general to a certain extent and should therefore be considered carefully in any applied context.

However, the study is limited in other ways as well. Firstly, the pure focus on the computational experiment confines the explanatory power as there are neither empirical real-world experiments nor the mathematical theorem considered. Additionally, due to computational limitations and for the sake of conciseness only a limited number of experiments with little to no variations in group sizes, correlation of the landscape, and maximum heuristic were conducted. This is different from previous research where exhaustive parameter searches have been conducted. However, given the knowledge gained from these prior contributions (Grim et al., 2018; Holman, 2018; Hong & Page, 2004; Singer, 2018) it has not been deemed necessary as the novel findings provided here aim at providing first evidence concerning efficiency as well as a confirmation of the anticipated effect of C-diversity. Also, the potential superiority of mixed groups has not been tested. Thus, these extensions and exhaustive searches could be addressed in future research. Beyond that, efficiency has only been quantified but no specific qualitative analysis of how groups reach the outcome has been conducted.

Conclusion

In conclusion, moving away from the mix-up of process of assembly and group properties as well as introducing efficiency to the debate revealed several insights. The distinction between the group properties has added further clarity as both appear to be mutually exclusive yet their combination is often the winning formula for final outcome with greater importance of one or the other depending on the landscape. This is supported by the literature, however, that a combination of both within one group is usually best could not be shown as explicitly beyond the testing of mixed groups (Grim et al., 2018). Moreover, to optimise efficiency, the complex interplay between diversity and expertise in a group needs to be taken into account. A combination of both is not always beneficial and practically difficult to realise due to the partially exclusive nature. The insights concerning efficiency are entirely new and need further research to be confirmed. Overall, there seems to be a strong contrast between diversity and ability, which can be seen based on the fact that the difference between clusters of groups which have been assembled to maximise one or the other often determine the results. Simultaneously, many groups can benefit from a combination of both presenting significant challenges.

These insights are valuable for a variety of reasons. First of all, they enhance the understanding of the theoretical model and its many components. Secondly, they add an initial consideration of efficiency. Thirdly, both these aspects together inform future research and could help practitioners make decisions, including anyone assembling teams. They could weigh the benefits and potential drawbacks of individual ability and diversity to compose better teams. However, the findings presented here are accompanied by a variety of limitations. Thus, further research is needed to draw definitive conclusions regarding efficiency and effectiveness. Future research could take the analysis further by focussing on multi-objective optimisation to find the optimal interplay between diversity and ability, potentially coupled with a more extensive parameter search including mixed groups. Beyond that, assembling more groups irrespective of any specific criteria may be a way to move away from the focus on specific group types to find new patterns determining performance. Furthermore, as pointed out earlier the monotonous tasks (only correlated at same strength or purely random) posed may expose groups to unlikely scenarios similar to many other modelling assumptions which could be remedied with or without combination with other models to paint a more realistic picture. Finally, empirical evidence can be gathered to test the hypotheses generated in this and other contributions, which would lend further credibility to the findings. While there is some evidence supporting the claim of diversity trumping ability regarding the final outcome (examples include Krause et al., 2011; Morand-Ferron & Quinn, 2011), there are many aspects on the several dimensions, which need to be tested next to the unknown territory of efficiency.

References

Agosto, D. E. (2002). Bounded rationality and satisficing in young people's Web-based decision making. *Journal of the American Society for Information Science and Technology*, *53*(1), 16–27. https://doi.org/10.1002/asi.10024

Alexander, J. M., Himmelreich, J., & Thompson, C. (2015). Epistemic landscapes, optimal search, and the division of cognitive labor. *Philosophy of Science*, *82*(3), 424–453. https://doi.org/DOI: 10.1086/681766

Alexandrova, A., & Northcott, R. (2013). It's just a feeling: why economic models do not explain. *Journal of Economic Methodology*, *20*(3), 262–267. https://doi.org/10.1080/1350178X.2013.828873

Bove, V., & Elia, L. (2017). Migration, diversity, and economic growth. *World Development*, *89*, 227–239. https://doi.org/https://doi.org/10.1016/j.worlddev.2016.08.012

Clearwater, S. H., Huberman, B. A., & Hogg, T. (1991). Cooperative solution of constraint satisfaction problems. *Science*, *254*(5035), 1181–1183. https://doi.org/10.1126/science.254.5035.1181

Grim, P., Bramson, A., Singer, D. J., Berger, W. J., Jung, J., & Page, S. E. (2020). Representation in models of epistemic democracy. *Episteme*, *17*(4), 498–518. https://doi.org/DOI: 10.1017/epi.2018.51

Grim, P., Singer, D. J., Bramson, A., Holman, B., McGeehan, S., & Berger, W. J. (2018). Diversity, ability, and expertise in epistemic communities. *Philosophy of Science*, *86*(1), 98–123. https://doi.org/10.1086/701070

Harris, C. R., Millman, K. J., der Walt, S. J. van, Gommers, R., Virtanen, P., David Cournapeau, Wieser, E., Taylor, J., Sebastian Berg, Smith, N. J., Kern, R., Hoyer, M. P. and S., van Kerkwijk,

M. H., Matthew Brett, Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Pierre Gérard-Marchant, … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Holman, B., Berger, W. J., Singer, D. J., Grim, P., & Bramson, A. (2018). Agent-based modeling in political philosophy. *Historical Social Research / Historische Sozialforschung*, *43*(1 (163)), 259–284. http://www.jstor.org/stable/26395052

Hong, L., & Page, S. E. (2001). Problem solving by heterogeneous agents. *Journal of Economic Theory*, *97*(1), 123–163. https://doi.org/https://doi.org/10.1006/jeth.2000.2709

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(46), 16385. https://doi.org/10.1073/pnas.0403723101

Katsikopoulos, K. v. (2014). Bounded rationality: the two cultures. *Journal of Economic Methodology*, *21*(4), 361–374. https://doi.org/10.1080/1350178X.2014.965908

Krause, S., James, R., Faria, J. J., Ruxton, G. D., & Krause, J. (2011). Swarm intelligence in humans: diversity can trump ability. *Animal Behaviour*, *81*(5), 941–948. https://doi.org/https://doi.org/10.1016/j.anbehav.2010.12.018

Landemore, H. (2021). An epistemic argument for democracy. In M. Hannon & J. Ridder (Eds.), *The routledge handbook of political epistemology* (pp. 363–373). Routledge.

Lijphart, A. (2012). Patterns of democracy: Government forms and performance in thirty-six countries. Yale University Press.

Mayo-Wilson, C., Zollman, K. J. S., & Danks, D. (2011). The Independence Thesis: When individual and social epistemology diverge. *Philosophy of Science*, *78*(4), 653–677. https://doi.org/10.1086/661777

Morand-Ferron, J., & Quinn, J. L. (2011). Larger groups of passerines are more efficient problem solvers in the wild. *Proceedings of the National Academy of Sciences*, *108*(38), 15898–15903. https://doi.org/10.1073/pnas.1111560108

Niazi, M., & Hussain, A. (2011). Agent-based computing from multi-agent systems to agent-based models: a visual survey. *Scientometrics*, *89*(2), 479. https://doi.org/10.1007/s11192-011-0468-9

Nisbett, R. E., & Ross, L. (1980). *Human inference: strategies and shortcomings of social judgement*. Prentice-Hall, Inc.

Macleod, W. B. (1996). Decision, contract, and emotion: Some economics for a complex and confusing world. *The Canadian Journal of Economics / Revue Canadienne d'Economique*, *29*(4), 788–810. https://doi.org/10.2307/136215

Polzer, J. T., Milton, L. P., & Swarm, W. B. (2002). Capitalizing on diversity: Interpersonal congruence in small work groups. *Administrative Science Quarterly*, *47*(2), 296–324. https://doi.org/10.2307/3094807

Sakai, R. (2020). Mathematical models and robustness analysis in epistemic democracy: A systematic review of Diversity Trumps Ability Theorem models. *Philosophy of the Social Sciences*, *50*(3), 195–214. https://doi.org/10.1177/0048393120917635

Singer, D. J. (2018). Diversity, not randomness, trumps ability. *Philosophy of Science*, *86*(1), 178–191. https://doi.org/10.1086/701074

Thoma, J. (2015). The epistemic division of labor revisited. *Philosophy of Science*, *82*(3), 454–472. https://doi.org/10.1086/681768

Thomas, D. A., & Ely, R. J. (1996). Making differences matter. *Harvard Business Review*, *74*(5), 1–13.

Thompson, A. (2014). Does diversity trump ability? *Notices of the AMS*, *61*(9), 1024–1030.

van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal*, *36*(3), 590–602. https://doi.org/10.5465/256593

Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, *76*(2), 225–252. https://doi.org/10.1086/644786

Weymark, J. A. (2015). Cognitive diversity, binary decisions, and epistemic democracy. Episteme, 12(4), 497–511. https://doi.org/DOI: 10.1017/epi.2015.34

Zollman, K. J. S. (2007). The communication structure of epistemic communities. *Philosophy of Science*, *74*(5), 574–587. https://doi.org/10.1086/525605

Zollman, K. J. S. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, *72*(1), 17. https://doi.org/10.1007/s10670-009-9194-6

Appendix A

Regression analysis of diversity and performance indicators

The following regression tables contain the results of the regressions with the diversity score (independent variable) and the performance indicator (dependent variable). Each regression is run with both diversity measures to score the groups and for both modes of collaboration. Each table shows the results for either correlated or uncorrelated and for groups whose maximum heuristics is either twelve or 20. Additionally, standard errors are reported in brackets. For further information please see the Jupyter notebook, details are in Appendix C.

Uncorrelated Landscape, Max. Heuristic: 12

	outcome, C_score, sequential	outcome, C_score, non-sequential	outcome, HP_score, sequential	outcome, HP_score, non-sequential	checks, C_score, sequential	checks, C_score, non-sequential	checks, HP_score, sequential	checks, HP_score, non-sequential	moves, C_score, sequential	moves, C_score, non-sequential	moves, HP_score, sequential	moves, HP_score, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
C_score	6.037***	3.613***			-0.052***	0.009			-14.358***	0.961***		
	(0.207)	(0.166)			(0.006)	(0.006)			(0.458)	(0.045)		
HP_score			12.670***	7.730***			-0.084***	0.037**			-31.556***	2.100***
			(0.997)	(0.665)			(0.018)	(0.017)			(2.216)	(0.175)
Intercept	89.251***	93.194***	82.916***	89.267***	1.577***	1.162***	1.608***	1.135***	53.448***	5.835***	69.830***	4.751***
	(0.188)	(0.151)	(0.923)	(0.615)	(0.005)	(0.006)	(0.017)	(0.016)	(0.415)	(0.040)	(2.051)	(0.162)
Observations	200	200	200	200	200	200	200	200	200	200	200	200
R^2	0.811	0.705	0.449	0.406	0.297	0.010	0.097	0.023	0.833	0.702	0.506	0.421
Adjusted R^2	0.810	0.703	0.447	0.403	0.293	0.005	0.093	0.018	0.832	0.700	0.504	0.419

Residual Std. Error	0.504(df = 198)	0.404(df = 198)	0.859(df = 198)	0.573(df = 198)	0.014(df = 198)	0.015(df = 198)	0.016(df = 198)	0.015(df = 198)	1.112(df = 198)	0.108(df = 198)	1.909(df = 198)	0.151(df = 198)
F Statistic	848.171*** (df = 1.0; 198.0)	472.610*** (df = 1.0; 198.0)	161.589*** (df = 1.0; 198.0)	135.273*** (df = 1.0; 198.0)	83.527*** (df = 1.0; 198.0)	1.938 (df = 1.0; 198.0)	21.312*** (df = 1.0; 198.0)	4.584** (df = 1.0; 198.0)	984.095*** (df = 1.0; 198.0)	466.376*** (df = 1.0; 198.0)	202.854*** (df = 1.0; 198.0)	144.247*** (df = 1.0; 198.0)

Note: *p<0.1; **p<0.05; ***p<0.01

Uncorrelated Landscape, Max. Heuristic: 20

	outcome, C_score, sequential	outcome, C_score, non-sequential	outcome, HP_score, sequential	outcome, HP_score, non-sequential	checks, C_score, sequential	checks, C_score, non-sequential	checks, HP_score, sequential	checks, HP_score, non-sequential	moves, C_score, sequential	moves, C_score, non-sequential	moves, HP_score, sequential	moves, HP_score, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
C_score	6.171***	3.480***			-0.104***	-0.013***			-15.440***	0.895***		
	(0.199)	(0.147)			(0.006)	(0.004)			(0.381)	(0.035)		
HP_score			21.819***	12.700***			-0.333***	-0.034**			-53.265***	3.127***
			(0.884)	(0.568)			(0.027)	(0.016)			(2.054)	(0.153)
Intercept	91.271***	94.781***	75.277***	85.385***	1.579***	1.159***	1.816***	1.182***	49.124***	6.005***	87.877***	3.721***
	(0.161)	(0.119)	(0.844)	(0.543)	(0.005)	(0.004)	(0.026)	(0.016)	(0.308)	(0.029)	(1.961)	(0.146)
Observations	200	200	200	200	200	200	200	200	200	200	200	200
R^2	0.830	0.739	0.755	0.716	0.586	0.043	0.437	0.022	0.892	0.764	0.773	0.679
Adjusted R^2	0.829	0.738	0.753	0.715	0.584	0.038	0.434	0.017	0.892	0.763	0.771	0.677

Residual Std. Error	0.620(df = 198)	0.458(df = 198)	0.744(df = 198)	0.478(df = 198)	0.019(df = 198)	0.014(df = 198)	0.023(df = 198)	0.014(df = 198)	1.188(df = 198)	0.110(df = 198)	1.727(df = 198)	0.129(df = 198)
F Statistic	963.909*** (df = 1.0; 198.0)	561.336*** (df = 1.0; 198.0)	608.744*** (df = 1.0; 198.0)	499.222*** (df = 1.0; 198.0)	280.270*** (df = 1.0; 198.0)	8.805*** (df = 1.0; 198.0)	153.692*** (df = 1.0; 198.0)	4.372** (df = 1.0; 198.0)	1641.754*** (df = 1.0; 198.0)	642.128*** (df = 1.0; 198.0)	672.690*** (df = 1.0; 198.0)	417.904*** (df = 1.0; 198.0)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Correlated Landscape, Max. Heuristic: 12

	outcome_corr, C_score, sequential	outcome_corr, C_score, non-sequential	outcome_corr, HP_score, sequential	outcome_corr, HP_score, non-sequential	checks_corr, C_score, sequential	checks_corr, C_score, non-sequential	checks_corr, HP_score, sequential	checks_corr, HP_score, non-sequential	moves_corr, C_score, sequential	moves_corr, C_score, non-sequential	moves_corr, HP_score, sequential	moves_corr, HP_score, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
C_score	0.101	0.307			0.006	0.175***			3.415**	2.096***		
	(0.318)	(0.327)			(0.007)	(0.009)			(1.458)	(0.104)		
HP_score			0.114	0.498			0.009	0.281***			6.670**	3.445***
			(0.562)	(0.578)			(0.013)	(0.018)			(2.567)	(0.208)
Intercept	54.053***	54.886***	54.039***	54.710***	0.983***	0.591***	0.980***	0.493***	27.715***	2.039***	24.739***	0.792***
	(0.284)	(0.293)	(0.506)	(0.521)	(0.006)	(0.008)	(0.012)	(0.016)	(1.304)	(0.093)	(2.311)	(0.187)
Observations	200	200	200	200	200	200	200	200	200	200	200	200
R^2	0.001	0.004	0.000	0.004	0.003	0.659	0.003	0.543	0.027	0.671	0.033	0.581
Adjusted R^2	-0.005	-0.001	-0.005	-0.001	-0.002	0.657	-0.002	0.541	0.022	0.670	0.028	0.579
Residual Std. Error	0.856(df = 198)	0.881(df = 198)	0.856(df = 198)	0.881(df = 198)	0.020(df = 198)	0.024(df = 198)	0.020(df = 198)	0.028(df = 198)	3.925(df = 198)	0.281(df = 198)	3.913(df = 198)	0.317(df = 198)

F Statistic	0.100 (df = 1.0; 198.0)	0.882 (df = 1.0; 198.0)	0.041 (df = 1.0; 198.0)	0.741 (df = 1.0; 198.0)	0.635 (df = 1.0; 198.0)	382.872*** (df = 1.0; 198.0)	0.534 (df = 1.0; 198.0)	235.293*** (df = 1.0; 198.0)	5.491** (df = 1.0; 198.0)	404.359*** (df = 1.0; 198.0)	6.752** (df = 1.0; 198.0)	274.659*** (df = 1.0; 198.0)
Note:									12, corr, *p<0.1; **p<0.05; ***p<0.01			

Correlated Landscape, Max. Heuristic: 12

	outcome_corr, C_score, sequential	outcome_corr, C_score, non-sequential	outcome_corr, HP_score, sequential	outcome_corr, HP_score, non-sequential	checks_corr, C_score, sequential	checks_corr, C_score, non-sequential	checks_corr, HP_score, sequential	checks_corr, HP_score, non-sequential	moves_corr, C_score, sequential	moves_corr, C_score, non-sequential	moves_corr, HP_score, sequential	moves_corr, HP_score, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
C_score	0.225	0.632**			-0.020***	0.119***			-0.483	1.669***		
	(0.258)	(0.261)			(0.006)	(0.008)			(1.075)	(0.097)		
HP_score			0.551	1.484***			-0.037***	0.250***			-0.960	3.467***
			(0.541)	(0.547)			(0.013)	(0.017)			(2.257)	(0.207)
Intercept	55.945***	57.082***	55.610***	56.200***	0.984***	0.634***	1.003***	0.496***	28.518***	2.659***	29.032***	0.745***
	(0.207)	(0.209)	(0.502)	(0.507)	(0.005)	(0.007)	(0.013)	(0.016)	(0.861)	(0.078)	(2.092)	(0.191)
Observations	200	200	200	200	200	200	200	200	200	200	200	200
R^2	0.004	0.029	0.005	0.036	0.045	0.518	0.036	0.519	0.001	0.600	0.001	0.587
Adjusted R^2	-0.001	0.024	0.000	0.031	0.041	0.515	0.031	0.516	-0.004	0.598	-0.004	0.585
Residual Std. Error	0.868(df = 198)	0.879(df = 198)	0.867(df = 198)	0.876(df = 198)	0.022(df = 198)	0.027(df = 198)	0.022(df = 198)	0.027(df = 198)	3.615(df = 198)	0.326(df = 198)	3.615(df = 198)	0.331(df = 198)
F Statistic	0.764 (df =	5.856** (df =	1.036 (df =	7.369*** (df	9.427*** (df	212.376*** (df	7.383*** (df	213.377*** (df	0.202 (df =	296.677*** (df	0.181 (df =	281.805*** (df =

1.0; 198.0)	1.0; 198.0)	1.0; 198.0)	= 1.0; 198.0)	= 1.0; 198.0)	= 1.0; 198.0)	= 1.0; 198.0)	= 1.0; 198.0)	1.0; 198.0)	= 1.0; 198.0)	1.0; 198.0)	1.0; 198.0)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Appendix B

Regression analysis of group types and performance indicators - absolute efficiency

The following regression table is an example of the regression tables used for research question one with one adjustment: here I use absolute numbers of checks and moves for efficiency. This specific table is a regression analysis of the different types of groups with maximum heuristic of twelve on uncorrelated landscapes. Each variable is a dummy variable for each type of group. The reference group type (intercept) is the group type of individually best-performing agents. The C coefficients are the coefficients for maximally C-diverse dummy variables similar to the HP coefficient, which is the coefficient for the maximally HP-diverse group type dummy variables. The columns represent the different performance indicators coupled with a specific mode of collaboration. Standard errors are reported in parentheses.

The table demonstrates the impact of the modes of collaboration on the efficiency in absolute numbers as described in the discussion section. Please consult the results section - first research question for the tables containing the results on relative efficiency. While this is only one example, the association between modes of collaboration and absolute number of checks and moves relative to each other holds across different heuristic spaces and types of landscapes. For further details please see the project notebook (https://doi.org/10.17605/OSF.IO/4YPK3).

	outcome, sequential	outcome, non-sequential	checks, sequential	checks, non-sequential	moves, sequential	moves, non-sequential
	(1)	(2)	(3)	(4)	(5)	(6)
C	2.172***	1.243***	2.157***	0.721***	0.324***	-0.674***
	(0.145)	(0.103)	(0.166)	(0.231)	(0.020)	(0.044)
HP	2.121***	1.318***	1.865***	0.352	0.319***	-0.660***
	(0.145)	(0.103)	(0.166)	(0.231)	(0.020)	(0.044)
random	1.927***	1.132***	1.870***	0.412*	0.283***	-0.681***
	(0.145)	(0.103)	(0.166)	(0.231)	(0.020)	(0.044)
Intercept (best)	93.067***	95.485***	60.375***	82.076***	2.106***	14.924***
	(0.102)	(0.073)	(0.117)	(0.163)	(0.014)	(0.031)
Observations	200	200	200	200	200	200
R^2	0.614	0.528	0.523	0.048	0.652	0.641
Adjusted R^2	0.608	0.521	0.516	0.033	0.647	0.635
Residual Std. Error	0.723(df = 196)	0.513(df = 196)	0.828(df = 196)	1.154(df = 196)	0.099(df = 196)	0.220(df = 196)
F Statistic	103.863*** (df = 3.0; 196.0)	73.064*** (df = 3.0; 196.0)	71.743*** (df = 3.0; 196.0)	3.287** (df = 3.0; 196.0)	122.641*** (df = 3.0; 196.0)	116.626*** (df = 3.0; 196.0)

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix C

Project notebook and datasets

The Jupyter notebook containing all steps of the simulation and the analysis of the data can be found here: https://doi.org/10.17605/OSF.IO/4YPK3. Furthermore, the notebook contains the plots on the assumptions for the regression analysis performed for research question one. Additionally, regression tables concerning the other research questions are included together with the figures shown here as well as additional ones on absolute numbers of checks and moves for some of the research questions. The code is modularised and thus easy to apply to any desired parameter setting (correlated(various degrees)/uncorrelated landscapes, modes of collaboration, heuristic spaces and other parameters)

Besides, the datasets used for the analysis provided here are included as well. Thus, it is not necessary to run the simulation again for a replication or extension on the basis of already present data. Some file names used in the notebook may differ, however, the ones uploaded contain all the data necessary under the exact same column names.