Modeling Direct & Indirect Effects of Cultural Diversity

Rik Warren
711 HPW/RHXB
U.S. Air Force Research Laboratory
Wright-Patterson AFB, Ohio 45433-7022, U.S.A.

Abstract

Cultural diversity can have both positive and negative effects on the performance of multinational teams. Positive effects can stem from the wide variety of experiences and methods that a culturally diverse team can draw on to plan and to develop innovative or efficient solutions to a complex problem. However, cultural diversity can also lead to miscommunications and misunderstandings during task execution. In a complex task involving separate planning and execution phases, it is hypothesized that cultural diversity promotes better strategy and richer ideas arising from the interaction of the team as a whole during the planning phase. Further, it is hypothesized that potentially positive effects from good planning can be partially countered in the execution phase since cultural diversity can lead to poorer communications and misunderstandings which arise from the many separate one-on-one conversations which take place during post-planning execution. This paper examines both direct and indirect effects of diversity. It also proposes a model incorporating competing mechanisms in order to understand performance in culturally diverse teams. Challenges to agent-based simulation in order to instantiate the model are discussed. These include the indexing of cultural diversity, and the simulation of planning and agent-communication effects. Results of the simulation are compared to data from human multinational teams engaged in a complex task.

1 Introduction

"To someone with only a hammer, everything looks like a nail." Since I am an experimental psychologist, my approach to testing an agent-based model of the effects of cultural diversity on team performance led to some unexpected problems. In turn, some problems have led to interesting solutions, and some still await answers.

Agent-based modeling (ABM)—with its focus on interacting entities which can be given their own rules of behavior and "personal" characteristics—is a "natural" tool for studying the effects of culture on group dynamics. But, as with other tools, we must learn its capabilities, limitations, and how best to use it effectively. In order to better utilize agent-based modeling, this paper addresses:

- Select difficulties and nuances of doing research on culture and group dynamics using human subjects which bear on ABM.
- Select unique capabilities and residual difficulties of using agent-based modeling.

My selections are not intended to be exhaustive but rather are offered as lessons-learned and lessons-waiting-to-be-learned. They are motivated by experiences with a large experiment on team performance using multi-cultural groups [1] and with developing an agent-based model of the surprising human results. Hence, this paper:

- Reviews the experiment with a focus on inherent human-based issues which might be eliminated using agent-based techniques.
- Discusses the development of an alternate hypothesis inspired by agent-based modeling.
- Reviews aspects of the agent-based model and simulation and its implementation.
- Discuses issues in testing the model and contrasts how testing a hypothesis with an agent-based simulation differs from testing the "same" hypothesis with a human-based experiment.

Agent-based modeling is known for revealing emergent properties and other "bonuses." One unexpected bonus of value for the study of cultural effects on group dynamics concerns the development of a metric for group diversity as will be highlighted below.

2 Human-based NATO Experiment

In order to investigate the performance of mixed- versus homogeneous-culture military teams, the NATO Human Factors and Medicine Panel on "Adaptability in Multinational Coalitions" (HFM–138) conducted a large five-nation computer-game based search-task experiment [1]. The search task required planning, resource allocation, situation awareness, communications and coordination for good performance.

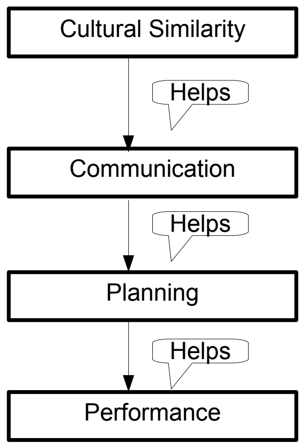

Figure 1: Implicit single-factor model of cultural effects.

2.1 Hypothesis & Rationale

Good planning, resource allocation, and situational awareness depend on good communication and coordination. In turn, good communication and coordination are facilitated by sharing a common culture. This logic chain (Fig. 1) assumes that culture is efficacious only via the single mechanism of communication. Hence, the principal hypothesis was: Homogeneous-culture teams (i.e., teams whose members are all from the same nation) perform better than mixed-culture teams (i.e., teams whose members are from different nations).

2.2 METHOD, NATO Experiment, Essentials

Details of the experiment are in [1,2]. Features most relevant to the agent-based simulation are:

2.2.1 Participants, Teams, & Unavoidable Demographic confounds

There were 56 teams of 4 NATO officers each: 8 from Bulgaria, 8 from The Netherlands, 16 from Norway, 9 from Sweden, 7 from the United States, and 8 with mixed national compositions. Eight of the Norwegian teams consisted of junior officers or cadets; the 8 other teams were more senior. Hence, some analyses treat these as two separate "culture" groups: No.j and No.s for "junior" and "senior."

No age requirements were set although an imposed similarity of ranks acted to keep ages

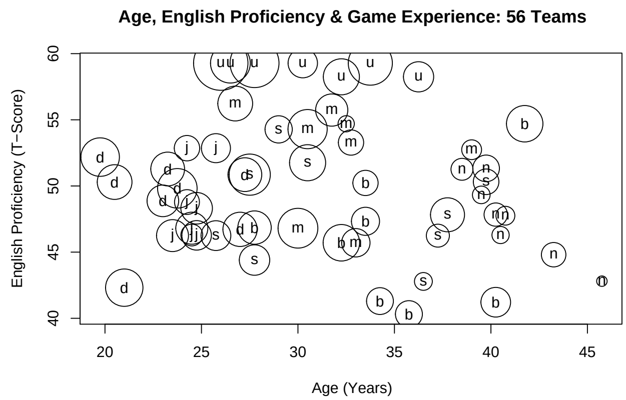

Figure 2: Demographic profiles of the 56 teams: Age, English proficiency, and computer-game experience. Game experience is proportional to size of bubbles. Letters indicate national composition of the teams. Key: Bulgaria (b), The Netherlands (d), Norway-senior age (n), Norway-junior age (j), Sweden (s), & the United States (u), Mixed culture (m).

within a team somewhat similar. No requirements were set for computer-game experience nor was it controlled for. All had to speak and write English, but no specific proficiency criterion was set. The result of these selection constraints is that age, English proficiency, and computer-game experience were not independent of each other or national composition.

These factors could not be experimentally controlled due to severe constraints on the available subject pool and the fact that the factors could only be measured *after* teams were formed and subjects present for the experiment. However, as mentioned, their effects can be assessed and partialled-out using linear regression techniques [2].

Fig. 2 is a bubble chart of the three demographic factors with the national composition of each team indicated. Distinct non-balanced non-factorially-crossed patterns are evident: For example, all seven American teams form a cluster at the high end of English proficiency and at the middle of the age scale. The bubbles indicate that the Americans also have relatively high levels of computer-game experience. The Dutch teams form another cluster at the younger end of the age scale and also show high levels of computer-game experience. The senior Norwegian teams, in contrast, form a cluster at the upper end of the age scale and show low levels of computer-game experience.

2.2.2 The Computer Game & Scenario

BBN Technologies, Inc. built a powerful system for research on culture around the immersive and absorbing role-play computer game *NeverWinter Nights*^{TM} [3–5], and developed a modern urban search-for-contraband scenario specifically tailored for this NATO experiment [6] which required planning, resource allocation, situation awareness, communication, and coordination for good performance. Good performance also required maintaining the good-will of the local populace who could provide useful tips or, the opposite, misinformation to the searchers. Essentially, teams were to find contraband caches hidden in a modern urban environment.

2.2.3 Procedure

Each team member was seated at a shielded computer terminal. Keyboards and computer screens were the only means of communication and information sharing with all communication in English. Mixed-nation team members were always in their home nation and played the game over the Internet.

The game is complex with three phases to the game-play:

- Training: Team-members received two-hours of training.
- Planning: Prior to the search phase, teams could study a map and were free to form their own search procedures. Team members could be specialists, e.g., communications officers, coordinators, weapon sensor carriers.
- Execution or search phase: There was a penalty for opening a suspected weapons cache if it was, in fact, empty. Hence, it was wise for a player without a special sensor to call for help from someone with a sensor.

2.2.4 Design & Performance Metrics

The primary independent variable was the homogeneous- versus mixed-culture composition of the 56 teams. The primary dependent variable was a *team* composite "goodwill" score.

2.3 Results

Contrary to expectations, homogeneous-culture teams were not generally better than mixed-culture teams (Fig. 3). In fact, after statistically removing the effects of several unavoidable confounding factors using multiple regression techniques, Warren [2] found that the mixed-culture teams were clearly superior (Fig. 4). Also noteworthy is that the raw data shows considerable differences between homogeneous teams from different cultures (Fig. 3), but the confound-removed data shows fewer differences between homogeneous teams (Fig. 4). The empirical study made no predictions about performance differences between cultures, however the empirical data justify the modeling effort's focus on just mixed- versus same-culture performance differences once we assume that all other factors are equal.

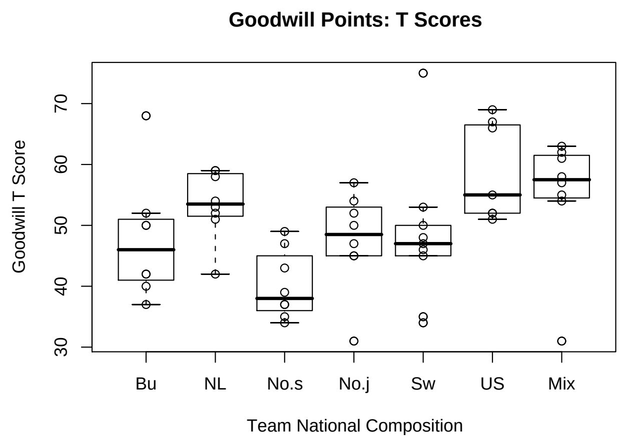

Figure 3: Overall game-play performance T-score (i.e., Mean = 50, SD = 10) for each of 56 teams grouped by national composition. Key: Bulgaria (Bu), The Netherlands (NL), Norway-senior age (No.s), Norway-junior age (No.j), Sweden (Sw), & the United States (US), Mixed culture (Mix). Box plots superposed on culture groups.

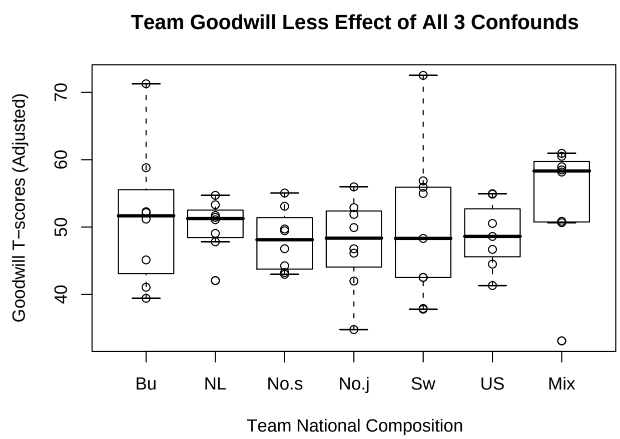

Figure 4: Game-play performance less effects of all 3 confounds (Adjusted T-scores: Mean = 50, SD = 7.87). Compare with Fig. 3.

2.4 Discussion

The chief experimental result is that the hypothesis that homogeneous-culture teams would perform better than mixed-culture teams was not supported. This surprise finding motivated an alternative hypothesis and the agent-based model discussed in the next section.

But there are other striking features that bear on modeling and empirical research on culture:

2.4.1 Participant Selection & Demographics

Cultural differences aside, human participants are not identical. Since individual differences are significant and pervasive, modeling and experimental research on culture are radically different than in the physical and engineering sciences.

The demographic "confounds" of age, gaming-experience, and English proficiency are unavoidable. Many differences cannot be eliminated by pre-selection, matched, or counterbalanced in an experiment since some demographic values are not measurable until after a person agrees to be in an experiment and provides data. Pre-selection is especially difficult, if not impossible, when a team is involved. For example, it is not uncommon to schedule six people when four are needed—and then to have only three show up for the experiment.

Yet the effects of the confounds are not insignificant. In the NATO experiment, the three confounds of age, gaming-experience, and English proficiency together accounted for 40% of the variance in performance scores. It appears that the best that can be done is for experimenters to be aware of confounds, measure them, and remove their effects by such methods as Analysis of Covariance (ANCOVA) or multiple-regression. If this step is not taken, the "raw" results can be seriously distorted and misleading. That is, apparent differences due to culture might, more easily, be explained by a confound or covariate.

2.4.2 Training

Unless an experiment is long-term (here left undefined), it is unlikely that training would be sufficient to overcome the effects of the demographic variables such as age differences in performance. Also, if the task is complex and difficult, it is unlikely that training would lead to equivalent ability to perform a complex task (outside of differences hypothesized as due to culture or cultural interaction).

3 Two-Opposing Factors Model of Cultural Effects

In addition to asking what the effects of cultural diversity on team-performance are, we can also ask about how—and when—these effects come about. In particular, after considering possible theoretical explanations for the superior performance of culturally-mixed teams, Warren [2] speculated that the diversity of the mixed-culture teams permitted better planning and produced a better search strategy in the sense of the *Wisdom of Crowds* [7].

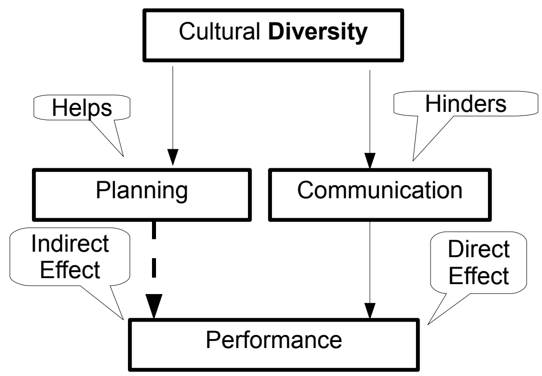

Figure 5: Two-factor opposing-mechanism model of cultural-diversity effects on a complex team task.

3.1 The Factors by Which Diversity Might Operate

This presumptive facilitating effect of diversity on search strategy clashes with the presumptive facilitating effect of homogeneity on communication and coordination. Moreover, whatever the relative "strengths" of the opposing factors, the possible mechanisms by which the putative factors might operate are not symmetric:

- The effects of diversity are arguably due to the interaction of the team *as a whole* in a planning phase, whereas
- the consequences of the quality of communications arise from the many individual one-on-one conversations which take place in the post-planning execution phase.

Under this analysis, culture has its effects via two mechanisms. Further, these mechanisms are presumed to be in opposition: as one promotes, the other hinders success (and *vice versa*). And still further, these mechanisms are presumed to operate in different phases. Two example scenarios help illustrate the extremes:

- A well-coordinated team might attempt to execute a poor plan.
- A team which develops a great plan might bungle its execution.

Fig. 5 illustrates this two-factor opposing-mechanism model of how cultural-diversity affects team performance.

3.2 Domain of Applicability & Complexity Considerations

The domain of applicability of the model thus assumes that the team's task spans two separate phases, namely, a planning phase and an execution phase. But a task that requires planning prior to execution suggests that a certain high-level of complexity is involved. Complexity here does not refer to the number of interacting agents, but rather to the nature of the interactions and their consequences. These interactions can be quite different in the various phases:

- Training: A task complex enough to require extensive training raises the question of training effectiveness. When homogeneous teams come from diverse cultures or when a team itself is culturally diverse, there is no guarantee that all teams are equivalently trained even when the same procedures and criteria are used.

- Planning: The task must be complex enough that planning is not just a matter of choosing between a few well-defined procedural alternatives. In the NATO experiment, teams had to determine role assignments and responsibilities, allocate resources, determine how to conduct the search, and formulate policies for dealing with unplanned events.

- Execution: The task facing the team must be complex enough that it cannot be carried out by independent agents. Success must require communication, coordination, and the asking for and giving of assistance. In the complex urgent real-world situations to which this model is intended to apply, it is possible for calls for help to not be heard, or for a potential help-provider to not be able to immediately respond or even not to be able to respond at all.

- Task objectives: An appropriate task to which the model applies need not have clear, well-defined objectives. Since many complex real-worlds tasks are not well-defined, defining a realistic objective can be considered a pre-planning phase with which a team must concern itself. In fact, it can be argued that the more fuzzy or ill-defined the objectives, the greater the potential benefit from having a culturally-diverse team.

3.3 Diversity Considerations: Team Diversity Index

A model of the effects of team cultural diversity requires an index of its central concept.

We begin with culture. Culture is a set of characteristics shared, more or less, by a group of people. Many individual dimensions have been proposed as particularly salient and useful for contrasting and comparing groups. These include power distance (e.g., [8]), Analytic-Holistic thinking (e.g., [9]) and Individualism-Collectivism (e.g., [10]). It is sometimes useful to further differentiate some dimensions such as Institutional Collectivism versus In-group Collectivism [11].

For these, and many other, proposed cultural dimensions, researchers have developed various questionnaires, sets of rating scales, and other tests. These measuring "instruments" are usually given to *individuals* and scored first to determine an individual's position along the cultural dimension in question. The distribution of one group's scores can then be compared with the distribution of another group's scores.

What is important for assessing team diversity is that scores on some dimension, or multidimensional composite, are available for each individual member of a team. Assume, then, that we have individual scores on some cultural dimension for all members of a team. How should we combine the scores to form a team diversity index?

The simplest team diversity index would be to use the (sample) standard deviation (SD) of the individual culture-index values. If the values were all equal, the SD's would be zero indicating no cultural diversity—no matter what the team size—as expected.

But, a big problem is that many culture-measures use different ranges. Some use a 5-point rating scale while others use a 7-point or more scale. Further, the final score might be reported as the mean rating or as the *sum* of the ratings. Since the SD is dependent on the range, such a diversity index would complicate comparisons and would not have a ready interpretation.

There is another problem with the standard deviation as a diversity index even when all the scores fall in the same range (say 0 to 100): The SD for [0 0 100 100] is 57.74, that for [0 0 0 100 100 100] is 54.77, and that for [0 0 0 0 100 100 100 100] is 53.45 indicating that the SD varies with the size of a team although the cultural composition is, in a sense, the same (here, equal parts from the extremes).

A better metric is to use the fraction (or percent) that the team SD is of the *maximal possible SD* given the particular range of values and the team size:

$$\text{team diversity index} = \frac{\text{actual team SD}}{\text{maximum possible team SD}} \tag{1}$$

Advantages of percent of maximal possible SD as as team diversity index include:

- Avoiding the problem of the dependency of simple SD on the scale or range of the measuring instrument and the dependency on team size.
- Statements like "10% diverse" and "80% diverse" are intuitively grasped and permit ready comparisons.

There are, unfortunately, some disadvantages of this particular index:

- I have not yet found a general formula or procedure for determining maximal SD given a particular range and team size. For now, it must be determined case by case.
- Maximal SD does not necessarily have complete intuitive appeal. My first guess was that a team with culture scores of [0 33 67 100] would be maximally diverse—and I still find the result hard to accept.

For reference in the simulation that follows: A maximally diverse team of four has two people at each of the extremes. The sample SD for [0 0 100 100] is 57.74

3.4 General Considerations & Comments

A few points should be clarified or emphasized:

- Diversity has its *effect* in the execution phase (since *all* effects are made manifest in the execution phase), but it is presumed due to the discussions in its planning phase and the "policies" and procedures that the team adopts.
- It is explicitly assumed that any communication problems of culturally diverse teams have a negligible impact during the planning phase. It is the (presumed) greater richness of ideas put forth and considered that matters, and it is assumed that the participants share enough of a common language for the ideas to "come across" no matter how awkwardly phrased, heavily "accented," or haltingly expressed from the viewpoint of native-speakers of the common language.
- However, awkward phrases, heavy accents, and hesitations while someone searches for a word can have negative impacts during the execution phase. These types of mis-communications cover anything from mis-hearings to misunderstandings.

The team diversity index permits using naturally or randomly configured teams in evaluations, but it also permits *assigning* a pre-selected degree of diversity to an artificial team. This enables a powerful method for testing the model.

4 Agent-based simulation: Method

The agent-base simulation attempts to capture certain aspects of the NATO experiment in exact mimicry, for example, by using teams of four interacting agents conducting a search for an unknown number of items. Other aspects are deliberately different. These especially include assuming (or setting) equal otherwise unavoidable confounds due to individual human and team differences in age, game-experience, English proficiency, motivation, training achievement, and the like.

The most interesting aspects of the simulation are how the *effects* of separate planning and execution phases are implemented since the oppositional interaction of the two-factors is the focus of the study. The key here is, as just intimated, to focus on the *effects* of planning and communication during execution rather than on the *processes* themselves.

Before presenting details about the agents, the environment, and search procedures, a synopsis provides the "big picture:"

4.1 Synopsis: Simulated Hunt for Treasure

Similar to the human game-play, four agents search—more or less efficiently—a space for hidden treasure caches. If an agent finds a treasure chest in a cave, the agent needs to call for assistance else it cannot retrieve the treasure (a gem). Unfortunately, an agent's call for help might not always be heard or understood. The performance metric is the number of gems a team collects within a fixed number of time-periods.

The agents are identical in their characteristics except possibly along a single cultural dimension. Teams, however, are not necessarily identical and can vary in their degree of cultural diversity.

The two-factor model states that more diverse teams produce better plans. Here, the effect of a better plan translates to using a more efficient search pattern. The model also states that the more homogeneous teams have better communication during the execution phase. Here, better communication during the search translates to agents having a greater radius within which their calls for assistance can be "heard" or "understood."

4.2 Searchers & Related Concepts

Searchers are the simulated equivalents of humans. Concepts related to searchers include the teams, the culture index, and the diversity index.

- Searchers: There are no demographic or psychological individual differences among the searcher except possibly along a cultural dimension. Searchers can vary on some temporary or contextual states such as their speed and current state with respect to being able to respond to a call for help.
- Culture index: Each searcher is assigned a number from 0 to 100 which represents their "position" on an unnamed "generic" cultural dimension, cultural cluster, or cultural group. The "0" and "100" carry no significance other signifying the extremes. "50" indicates the middle of the dimension.
- Team: The team consists of four searchers. Unlike the individual searchers, there is no formal computer entity or agent declared as "team."

4.3 The Environment

The world is fairly simple. It is a 33×33 square grid of 1089 patches. Since any agent who travels beyond one of the sides of the square boundary reenters from the opposite side, the world is actually topologically a torus. The world contains 20 caves scattered randomly. Each cave contains a treasure chest which contains a gem.

4.4 A Diversity-Dependent Search Pattern

The search strategy adopted here is only one of many possible strategies. All team members use the same basic search strategy which is based on the notion of a corridor. Corridors depend on the size of a heading window which is the same for all team members. Once the heading window is computed, the individual paths of the team members may be computed.

4.4.1 Corridors and efficient searches

A maximally efficient search pattern (for one person) is a continuous path that traverses all areas of an environment without revisiting previously visited areas—or with minimal backtracking if forced by the layout or geometry. Less efficient search patterns produce unnecessary backtracking or revisiting of prior locations.

If an agent meanders randomly (changing its heading at every move), it will often reenter previously visited patches before it enters all or most patches. If an agent moves with a constant heading, i.e., in a perfectly straight line, in most directions, it will wind around the entire (toroidal) world with no or little overlap of previously visited patches. (Think of a wire wound in a not-too-loose not-too-tight spiral around a cylindrical or toroidal core. Some directions, such as due north or due east, force paths that form circular or too-loose spiral loops which miss most patches.)

In between the extremes of no change in heading (greatest efficiency search excepting those headings which produce a "rut") and a completely random (up to 360°) change in heading at every step, random changes in heading of less than 360° should produce an intermediate level of search efficiency. For an initial heading, h_0, random changes within a $\pm x^\circ$ window generally produce a somewhat zig-zagging path which *averages* to that produced by a constant h_0 heading path. That is, random heading changes within a given range produce a path whose envelope forms a corridor aligned with that of a constant h_0 heading path. The width of the corridor is a function of speed and the size of the circular-arc window. Also, the width of the corridor is *inversely* related to search efficiency since narrower corridors generally cross more patches with less backtracking whereas wider corridors permit more revisiting of previously traversed areas.

The efficiency of a search pattern is assumed to be a consequence of the quality of the planning a team "conducted" prior to the start of the search. In turn, the quality of planning is, according to the two-factor model, directly proportional to the cultural diversity of the team as a whole. That is, the more diverse a team, the better the search pattern it produces.

4.4.2 Team heading window

We can use the corridor approach to develop a search procedure which varies in effectiveness as a function of team diversity:

- Compute the team diversity index (D_{team}) per (1).
- Choose a minimum angular width ω_{min} for selecting random headings. A ω_{min} greater than zero reflects that it is unlikely that a team selects a perfect search strategy. It permits a little jitter to avoid the occasional ruts that a straight path might fall into.
- Choose a maximum angular width ω_{max}. It is unlikely that a team adopts a search strategy as inefficient as totally random wondering, so ω_{max} can be considerably less than 360°.

The heading window which is used throughout a search run is thus:

$$\text{heading.window} = \omega_{max} - D_{team}\,(\omega_{max} - \omega_{min}) \tag{2}$$

For a maximally diverse team, (2) reduces to ω_{min}, and for a non-diverse team, it reduces to ω_{max}. By *randomly* choosing a heading *adjustment* in this window at each search step, the more diverse teams will search using narrower and more efficient corridors than less diverse teams.

4.4.3 Implementation: Heading changes

This procedure is simple to implement in an agent-based simulation.

- Each team member, i, must be assigned its own general heading, $h_{0,i}$.
- At each search step, t, each individual team member's heading is given by:

$$h_{t,i} = h_{0,i} \pm (1/2) \times \text{random(heading.window)} \tag{3}$$

The general headings, $h_{0,i}$, and thus the alignment of the corridors, that each individual is assigned may all be different or all the same. There is a surprise lurking in this choice that will be presented in the simulation results section.

4.5 Diversity-Dependent Communications

Whereas all team members use the same basic search strategy which is dependent on the cultural diversity of the team as a whole, communication effectiveness here depends on pair-wise person-to-person cultural similarity. Also, instead of using search corridor widths to (inversely) vary search effectiveness, communication range is used to directly reflect communication quality. Otherwise, there are many similarities between the diversity-dependent communication techniques adopted here and the diversity-dependent search technique.

4.5.1 Communication range & communication quality

The quality of event-by-event communication is, according to the two-factor model, inversely dependent on the cultural diversity of the specific two people communicating during the execution phase of a search. By focusing on the effect of communication quality, we can ignore the specifics by which a particular communication might be affected such as by heavy accents, halting expression, or limited understanding of idioms or cultural-dependent references.

The main communication message by the agents is the call for assistance that an agent broadcasts when a treasure chest is found. Another agent can only lend assistance if it "hears" or "understands" the call. The key here is to substitute distance for communication quality. That is, an agent's call for assistance can be "heard" or correctly "understood" at a further distance by a someone who shares the caller's language, idioms, and accent, than by someone who is of a different language, etc.

Thus, only an agent who is "nearby" can answer a call for assistance, and "nearby" depends on the cultural similarity of the caller and potential responder.

4.5.2 Determining communication range

To determine the effective communication range between two agents:

- Divide the absolute difference in the culture indices of the agents by the maximum possible difference. This is a fraction between zero and one and indicates how culturally disparate the two agents are:

$$D_{pair} = \left(\frac{\text{Abs(pair.dissimilarity)}}{\text{max.dissimilarity}} \right) \tag{4}$$

- Choose, d_{min}, the minimum distance (1 patch width is the unit distance) within which all messages are accurately sent and received by every pair of agents.

- Choose, d_{max}, the maximum distance beyond which no messages are accurately sent and received by even the most similar possible pair of agents.

The effective communication radius between a calling agent and potential responder is thus:

$$\text{comm.radius} = d_{max} - D_{pair}\,(d_{max} - d_{min}) \tag{5}$$

For a maximally diverse *pair*, (5) reduces to d_{min}, and for a non-diverse team, it reduces to d_{max}.

After determining the heading window (2), a further step was needed to determine the heading change an agent would take, namely, randomly choosing a value within the heading window. No such extra step is needed in using (5).

4.5.3 Implementation: Communication radii

Unlike randomly choosing a value from a common and fixed range to determine specific heading adjustments, the value of a communication radius depends on the cultural difference of the particular caller and potential-responder pair. If there are n searchers on a team, then there are: $\frac{n(n-1)}{2}$ unique searcher pairs. Depending on the culture values owned by the agents, there can be fewer than $n(n-1)/2$ unique communication radii. For example, three agents with cultural values [5 5 5] generate three pairs all with zero pairwise disparity and thus all having the same minimal communication radii. But, three agents with cultural values [0 5 10] generate two pairs with the same communication radii since the pairs [0 5] and [5 10] both have the same pairwise cultural disparity.

Implementation is a little more complicated than choosing a random value.

- Each team member, i, must be assigned its own cultural-factor value.
- The search for "nearby" agents needs only be done when an agent issues a call for help.

4.6 Equations (2) & (5) compared

Equations (2) and (5) are remarkably similar in structure: They are both straight lines with a negative slope. How then can they generate opposite effects as a function of (team or pair) diversity as required by the two-factor model? In both cases, greater diversity yields a smaller output. In the case of (2), the smaller output is a smaller random-walk corridor which results in a more efficient search patter, whereas in the case of (5), the smaller output is a smaller communication radius which results in fewer answered calls for assistance.

4.7 Search Procedures & Simulation Details

The simulation was developed using *NetLogo* version 4.0.4 [12]. After a set-up phase, the search proceeds as follows:

- At the start of every run or trial, the entire team starts at the center of the environment as if they just finished a planning session. Until a prescribed number of "ticks" is reached, the team members engage in two primary activities: searching and communicating. There is also the technical activity of retrieving treasure.
- Initially, the agents fan out in different directions using the same "agreed upon" search plan and at the same speed. If an agent finds a cave, it must call for help to open the treasure chest in the cave and retrieve a gem.
- If the call for help is not answered, the caller returns to active search at the next "tick."

- If the call is answered by another agent, a gem is retrieved and both agents return to active search at the next tick.

5 Agent-based Simulation: Tests & Results

The main per-trial performance measure is the number of gems (out of a maximum possible of 20) that a team finds during a trial of length of 1,000 "ticks." All means are based on runs of 200 trials each. If moving, all searchers move at 1 unit distance per tick. All other simulation parameters are as given earlier or below.

Before presenting the results about the effects of cultural diversity, I first treat a technical issue:

5.1 Search Success & Individual Search Corridors

This simulation test was not performed to test the model directly but to confirm the technical assertion that narrower search corridors produce better searches. A second variable, individual heading, h_0, range was included to test a suspicion that emerged during simulation development. A third major factor's values must be set for a simulation to execute, namely, team cultural diversity.

5.1.1 Team cultural diversity

The four team member's individual positions along the single generic cultural dimension were set at [0 33 67 100]. That is, the team members' values included the opposite extremes and two equispaced values in the range. The SD for these four values is 43.12 which is 74.7% of the maximum possible SD of 57.74. This high level of team cultural diversity was chosen to enable any diversity effects to clearly emerge. Team cultural composition was held constant throughout the testing of this first simulation.

5.1.2 Heading-change ranges & corridor widths

Search corridors can vary from a totally random walk to a perfect straight line path. An actual corridor width depends on the maximum and minimum arc-window widths imposed by the experimenter and the cultural diversity of the team (see (2)). The minimum heading-change window was always 15° and the maximum was either 90, 180, or 360°. Actual window widths were 33.99, 56.77, and 102.33° indicating that this highly diverse team "adopted" a (presumably) efficient search technique of always moving generally forward but allowing for three trial-dependent levels of random heading changes.

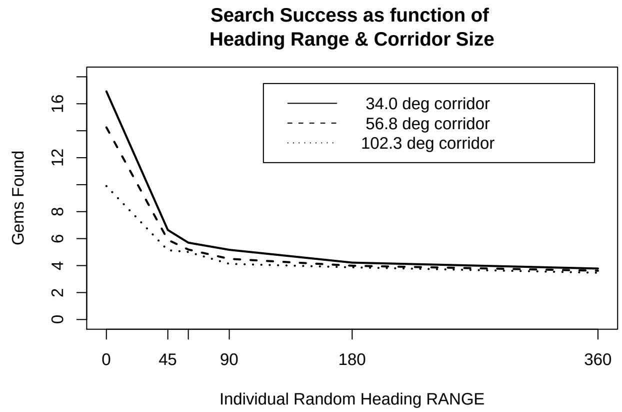

Figure 6: Search success as a function of the alignment range of individual search headings and size of the window for heading adjustments. All means based on 200 trials. Maximum possible gems found is 20.

5.1.3 Search directions & corridor alignments

Each team member moves somewhat randomly—zig-zagging more or less—in a corridor or lane whose width is determined as above and whose general direction, $h_{0,i}$, is (currently) set by the experimenter. Step-by-step actual headings are determined by (3).

The general headings, $h_{0,i}$, and thus the alignment of the corridors, that each individual is assigned may all be different or all the same. I initially assigned these totally randomly, but discovered during preliminary testing that search performance increased if all paths had the same direction and were parallel. Hence, for this test, six levels of path alignment were included: All general directions equal, or all chosen randomly from either 45, 60, 90, 180 or 360°.

5.1.4 Results & discussion: Corridor widths & alignments

Fig. 6 shows the results of the simulation test for search performance as a function of corridor alignment and width: Smaller search corridors produce better searches as expected. However, this holds true only for teams whose individual member general path headings are parallel or roughly aligned. If individual directions are chosen randomly over a 45° range, the negative effects of non-direction-alignment greatly reduces any possible positive effects of narrower search corridors. If individual directions are chosen randomly over a 180° range, the negative effects of non-direction-alignment completely eliminate any possible positive effects of narrower search corridors.

This powerful effect of direction or corridor alignment is the surprise I earlier said was lurking in (3). What I thought was a technical detail now appears to enable a powerful

Table 1: Team Cultural Composition & Diversity Index

Culture Values				SD	% of Max SD
50	50	50	50	0.00	0
0	15	30	45	19.36	33
0	30	60	90	38.73	67
0	0	100	100	57.74	100

possible alternative technique by which a team's planning effectiveness can be simulated. That is, teams can "adopt" path alignment values as well as corridor width values as a result of their "planning process."

5.2 Search Success & Cultural Diversity

This simulation test was performed to directly test the two-factor model's prediction that positive effects of diversity can counter problems arising from cultural-based problems of communication. As for the previous test, values must be set on three major factors: Diversity, corridor widths, and search directions.

5.2.1 Team cultural diversity

Four sets of team cultural composition were used to yield team cultural diversities of 0, 33, 67, and 100%. Table 1 shows the culture compositions, team SD's, and diversity values.

5.2.2 Heading-change ranges & corridor widths

The heading change window limits were fixed at 15° and 120°. This produced actual window sizes of 120°, 87.78°, 49.56°, and 15°, respectively for teams with 0, 33, 67, and 100% diversities. Step-by-step motion was generally forward with no or only moderate direction changes to encourage search success and avoid ceiling and floor effects.

5.2.3 Search directions & corridor alignments

The corridor test showed that search directions beyond a 180° random range can nullify heading-window width effects. Hence, three levels of search direction alignment ranges were included in this test but with none greater than 180°. The three levels of alignment were 0° (all path directions parallel), 90°, or 180°.

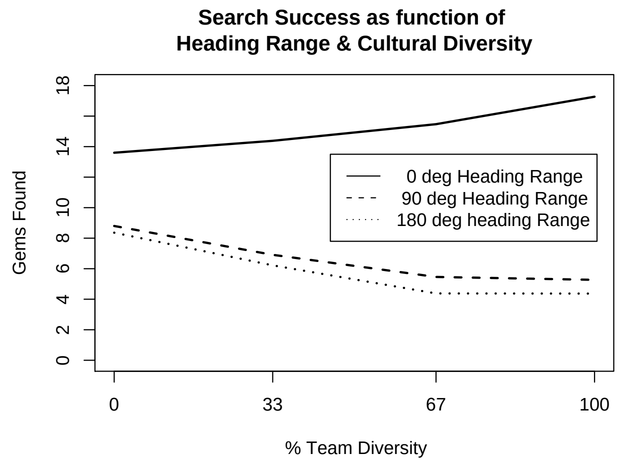

Figure 7: Search performance as a function of team-cultural diversity and search-path direction alignment.

5.2.4 Results & discussion: Diversity & direction alignments

Fig. 7 shows the results for search performance as a function of team-cultural diversity and search-path direction alignment: When search directions are parallel, search performance improves as a function of team cultural diversity. However, a path direction alignment range of 90° completely reverses the effect. Although, we know from the previous test that direction alignment is a powerful factor, the effect reversal is a surprise. Why should a less diverse team (which uses a wider, presumably less efficient search pattern) outperform a more diverse team (which uses a narrower presumably more efficient search pattern)?

5.2.5 A possible explanation of alignment effects

A possible answer to why a wide search pattern can result in superior performance when team members travel in different directions is that a wide pattern might increase the probability that an agent is nearby enough to answer another agent's call for assistance.

- If team members generally travel on parallel paths, then they might tend to travel as a "pack," more or less. This increases the chance that some member "hears" a call for assistance when the corridor widths are narrow (as "favored" by more diverse teams).
- If team members travel in scattered directions and in narrow corridors, they will cover a smaller area. But the smaller covered area means that some calls for assistance might not be heard.

- If team members travel in scattered directions and in broad corridors, they might cover a larger area which, in turn, might increase the likelihood that a call for assistance is heard.

Hence, a narrow search path might not always be the better technique as I originally thought. This has yet to be tested, but it might open up more possible strategies for a culturally diverse team to "consider."

6 General Discussion

The purpose of the current study was to use agent-based modeling to better understand the reasons behind the surprising results of the NATO–HFM–138 experiment, namely, that mixed-culture teams outperformed homogeneous-culture teams in a search task using a role-play computer game. The results of the agent simulation themselves held surprises. There are several aspects of the simulation study that are noteworthy concerning the simulation and its results, the two-factor model, the team-diversity index, and cultural research in general.

6.1 The Simulation

The simulation attempted to mimic certain aspects of the NATO–HFM–138 task, in particular using a search task requiring calls for help to team members. Mimicking *how* a team plans or communicates was not a goal. Rather, the simulation attempted to use techniques which would (re-)produce the effects of planning and communication quality.

Searching for hidden treasure met the mimicry requirement. Searching can be random or involve a strategy for determining where to move to. Strategy entails setting a basis for changing direction. The basis for changing direction was the technique for capturing the effects of planning quality. However, actual search requires that a speed (always a constant here) and a general direction also be set (per (3)) for each individual agent. As discussed earlier, the direction-setting procedure appears to be powerful, if not more important, than the direction-changing procedure. This was totally unexpected, but its discovery opens up new possibilities for simulating the effects of planning.

Many alternative techniques could be employed in lieu of the specific planning and communication functions used here. For example, a decaying exponential could be used for determining communication distance. What matters is that diversity can result in different action consequences, some beneficial and some not beneficial to performance.

The current simulation has not yet fully implemented features which would make the environment less benign such as traps and snares. Traps and snares would lend an urgency for calls for assistance, or at the least, promote some precautionary behaviors. Precautionary behaviors have some interesting cultural variations which would enrich the model.

In short, the simulation has a way to go toward providing for a more realistic complex world, but key elements have been captured.

6.2 The Two-Factor Model

The two-factor model is consistent with the simulation results. Fig. 7 shows that, under certain circumstances, team cultural diversity can produce superior performance in complex tasks which require separate phases for developing strategy and task execution. Fig. 7 also shows that, under different circumstances, team diversity can hinder performance. This is a far richer result and a richer explanation than the simple one-factor model depicted in Fig. 1, namely that cultural diversity can hamper team performance because of poor communication. The two-factor model depicted in Fig. 5 has a much richer role for culture to play allowing for both direct and indirect effects. Furthermore, the effects of cultural diversity need not always be influenced by communication problems.

The two-factor model is not without its limitations. It is currently featured as encompassing two opposing mechanisms with no alternatives. This is obviously unrealistic. For example, there are situations in which an awareness of cultural differences can promote an interest in different culture groups to learn more about each other and to communicate more for positive effects. Also, an awareness of cultural differences among groups engaged in a complex cooperative venture can engender—in the planning phase—a search for mitigation techniques (such as the use of codes or task-specific vocabulary) to be used in the execution phase.

In short, the model, like the simulation, has a way to go toward providing for a more realistic complex world, but key elements have been captured.

6.3 The Team-Diversity Index

The team-diversity index (1) is not yet in final form and has not been fully studied. However, it is clear that such an index is needed for strong experimental research on culture using either humans or agent-models. If we do not have or report such an index, then we will not be able to know how much team differences in performance are due to team differences in culture. Most measures and data on cultural factors are for individuals or for large national or ethnic groups. But how such individual or large-group values relate to the composition of a specific team is another matter.

One particular benefit of having a team-diversity index (relative to one or more measurable cultural factors), such as (1), with known end-points (i.e., 0% and 100%) and determinable intermediate values is that teams with specific values may be formed, for example, as in Table 1. Once this is done, either with humans or computer agents, powerful experiments may be designed which use known levels of team diversity. This is in sharp contrast to working with teams whose diversity values are not pre-selected. For example, if teams are formed at random, say from volunteers, it is possible that most teams' (unknown) diversity values might cluster at one part of the full range of possible values. This would then expose the research to statistical problems due to the restricted range.

6.4 Culture Research & Agent-Based Simulation

There is much interest today in cultural effects on cognition and performance. But research on culture and inter-cultural relations is extremely difficult to conduct and filled with unavoidable confounds which make interpretation of results difficult. Agent-based modeling and simulation enables us to test theories and uncover new and emergent phenomena so that we may may the most of precious & limited human subject pools.

References

[1] NATO HFM–138, Final Report of the NATO Human Factors and Medicine Panel Working Group 138 on "Adaptability in Multinational Coalitions." Brussels: NATO, 2008.

[2] R. Warren, "Mixed- & homogeneous-culture military team performance on a simulated mission: Effects of age, game-experience, & English proficiency," *Proceedings of the NATO RTO HFM–142 Symposium on Adaptability in Coalition Teamwork*, held in Copenhagen, Denmark. Brussels: NATO, 2008.

[3] Bioware Corp, *Neverwinter Nights*^{TM} - Platinum Edition (Computer Game Software). New York: Atari Interactive, 2004.

[4] R. Warren, J. Sutton, D. Diller, W. Ferguson, and A. Leung, "A game-based testbed for culture & personality research," *Proceedings of the NATO Modeling and Simulation Group – 037 Workshop: Exploiting Commercial Games for Military Use.* The Hague, The Netherlands, 20–21 Oct. 2004.

[5] A. Leung, D. Diller, and W. Ferguson, "SABRE: A game-based testbed for studying team behavior," *Proceedings of the Fall Simulation Interoperability Workshop (SISO).* Orlando, FL, September 18–23, 2005.

[6] R. Warren, D.E. Diller, A. Leung, W. Ferguson, and J.L. Sutton, "Simulating scenarios for research on culture & cognition using a commercial role-play game," In M.E. Kohl, N.M Steiger, F.B. Armstrong, and J.A. Jones, (Eds.), *Proceedings of the 2005 Winter Simulation Conference.* Orlando, FL.

[7] J. Surowiecki, *The wisdom of crowds.* New York: Anchor Books, 2004/2005.

[8] G. Hofstede, *Cultures consequences: Comparing values, behaviors, institutions, and organizations across nations,* (2nd ed.). Thousand Oaks, CA: Sage, 2001.

[9] R.E. Nisbett, *The geography of thought: How Asians and Westerners think differently . . . and why.* New York: The Free Press, 2003.

[10] H.C. Triandis, *Individualism and collectivism.* Boulder, CO: Westview Press, 1995.

[11] R.J. House, P.J. Hanges, M. Javidan, P.W. Dorfman, and V. Gupta, (Eds.), *Culture, leadership and organizations: The GLOBE study of 62 societies.* Thousand Oaks, CA: Sage, 2004.

[12] U. Wilensky, *NetLogo, Ver. 4.0.4.* http://ccl.northwestern.edu/netlogo. Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL, 1999/2008.