

More than our Differences:

The Effects of Anticipated Diversity on Cognitive Performance

A thesis submitted by

Satia Adele Marotta

in partial fulfillment of the requirements for the degree of

Master of Science

in

Psychology

Tufts University

February 2015

Adviser: Samuel R. Sommers, PhD

## Abstract

Research on the effect of diversity on performance suggests that in order to benefit from diversity, differences in knowledge, values, and social category must be synthesized. However, little is known about how our thinking may change when merely anticipating diversity in its simplest forms. This study assigned 132 participants to novel groups using the minimal group paradigm to ascertain potential differences in cognitive performance when participants anticipated interacting in either a homogeneous or diverse group. Results suggest that anticipating a diverse interaction may improve performance on some creativity and logical reasoning tasks, though findings were inconsistent across outcome measures. Anticipating diverse, as opposed to homogeneous, interactions may influence our cognitions about others and our performance on tasks in various and sometimes unanticipated ways.

## Table of Contents

|  |    |
|--|----|
| Abstract .....   | ii |
| More than our Differences: The Effects of Anticipated Diversity on Cognitive Performance ..... | 1  |
| Method .....   | 7  |
| Results.....   | 12 |
| Discussion.....  | 20 |
| References .....   | 26 |
| Table 1 .....  | 31 |
| Table 2 .....  | 32 |
| Appendix A.....  | 33 |

## More than our Differences: The Effects of Anticipated Diversity on Cognitive Performance

The race, ethnicity, and age demographics of the United States are rapidly becoming more diverse (Taylor, 2014). As many social policies continue to foster diversity in the workplace and higher education, we must investigate the influence these demographic changes will have on our experiences with others. Research has examined the influence that diversity can have on interpersonal relationships and group productivity, but results suggest that positive outcomes observed in creativity and problem-solving may be overwhelmed by negative outcomes observed in how groups collaborate and communicate. The present study investigated how anticipating a diverse interaction, when diversity is without historical or social context, can influence an individual's performance on cognitive tasks.

Social identity theory posits that we each maintain a social identity that is composed of our group memberships and our emotional attachments to those groups (Tajfel, 1974/2010). We tend to affiliate ourselves with the members of our groups, ingroup members, and align ourselves against individuals outside our groups, outgroup members. Often, this emotional attachment can lead us to try to enhance our self-image by improving our group's image. Sometimes this behavior can present as biases in favor of our ingroup members and against our outgroup members. While in some instances this ingroup favoritism has no harmful effect on other groups, when members of different groups must interact, the result can be an interaction that is taxing, hostile, or unproductive.

Increasing the frequency of intergroup interactions could have negative consequences for businesses and their employees. Diverse groups –whether diverse in age, gender, social category, tenure, or another criteria –can be costly as they tend to have higher instances of turnover and report more cases of health related issues and absences than homogeneous groups (DeDreu & Weingart, 2003; Williams and O'Reilly, 1998). Additionally, diverse groups tend to have difficulty synthesizing and implementing ideas while staying within budget and respecting deadlines (Harvey, 2013; Kerr &

Tinsdale, 2004; Mannix & Neale, 2005). These negative outcomes are exacerbated in businesses that are actively diversifying their workforce with programs such as affirmative action because this can alienate members of majority groups who may perceive these initiatives as discriminatory (Norton & Sommers, 2011; Outten, Schmitt, Miller, & Garcia, 2012).

The negative outcomes for businesses may be explained by the increased conflict often experienced between the members of diverse groups (DeDreu & Weingart, 2003). The members of homogeneous groups tend to have more frequent and productive communications between group members than members of diverse groups (Jackson & Joshi, 1999; Jehn, Northcraft, & Neale, 1999; Mannix & Neale, 2005). Additionally, members of diverse groups tend to communicate more with people outside of the group than with people within the group. In some instances, communication with external group members can provide a wider range of resources to the group; however, the positive aspect of this external communication is often under-utilized in diverse groups because they tend to refrain from discussing information that is not already known to all group members (Mannix & Neale, 2005).

Further, the composition of a diverse group can aggravate these negative outcomes; people who are accustomed to being in the majority tend to react more negatively to diverse interactions than those accustomed to being in the minority (Jackson & Joshi, 1999; Mannix & Neale, 2005; Phillips, Liljenquist, & Neale, 2009; van Knippenberg & Schippers, 2007; Williams & O'Reilly, 1998). The exception to this is when a color-blind strategy is used in the group's interactions; if differences are actively ignored majority group members tend to excel (Cunningham, 2009; Plaut, Thomas, & Goren, 2009). This occurs because while appearing egalitarian, color-blind strategies encourage sustaining the status quo for the group already in the majority as differences are not acknowledged or accepted, leading minority group members to feel unwanted.

However, not all diverse interactions are marked by negative outcomes; there are also benefits to diversity. Diversity within workgroups can lead to task conflict –disagreements about how to complete the task. Task conflict, in comparison to group or emotional conflict, can generate novel and creative ideas as different perspectives are integrated to create solutions that are pleasing to all group members (DeDreu & Weingart, 2003; McLeod, Lobel, & Cox, Jr., 1996; Williams & O'Reilly, 1998). Diverse groups also tend to have more varied insights, broader social networks, and greater idea exchange than homogeneous groups (Jackson & Joshi, 1999; van Knippenberg & Schippers, 2007). Additionally, within education, cross-cultural research suggests that as the percentage of immigrants at a school increases, there are corresponding improvements in the academic performance of both native and immigrant students (Konan, Chatard, Selimbegović, & Mugny, 2010).

There is little consensus on how to predict whether a diverse group will experience positive or negative outcomes. Some research suggests that there is a curvilinear relationship between diversity and performance; as the diversity of a group increases there is a corresponding increase in performance up until a point at which performance begins to suffer (DeDreu & Weingart, 2003; Williams & O'Reilly, 1998). This finding may be explained by how easily subgroups can form within a diverse group (Jackson & Joshi, 1999; van Knippenberg and Schippers, 2007). If subgroups are easily formed, whether by age, gender, race, tenure, or some other form of diversity, this can cause a fracture within the group preventing the cohesion and exchange of ideas that would result in positive outcomes.

Research is needed to determine how best to capitalize on increased diversity. Research on the circumstances in which positive outcomes can be observed in diverse interactions is mixed. Some researchers suggest that the benefits can only be observed if the diverse group is managed and monitored to aid in dispelling emotional conflicts between group members while encouraging task

conflicts that can lead to novel solutions (DeDreu & Weingart, 2003; Jehn, et al., 1999; Williams & O'Reilly, 1998). Other strategies to improve positive outcomes involve explicitly stating the workgroup's goals while minimizing interdependence between group members (Jackson and Joshi, 1999; Mannix & Neale, 2005). This encourages group members to collaborate in brainstorming while minimizing their reliance on others to complete their tasks. Additionally, increasing communication between group members and the time that the group has to work together can increase their feelings of similarity and improve their relationships. However, while these strategies might be practical in the workplace, in academic settings where students collaborate on projects for short periods of time these strategies may not be as effective.

Instead, some research has focused on the aspects of diversity that might contribute to improved performance in certain domains. Demographic or social category differences may be perceived as a cue to psychological differences that can provide varied perspectives (Mannix & Neale, 2005). Previously, the benefits associated with diversity were attributed to the novel perspectives and insights that minorities can provide as a result of these implied psychological differences (Antonio, Chang, Hakuta, Kenny, Levin, & Milem, 2004). However, this perspective places an undue burden on minorities to provide "the minority opinion," when in reality minority experiences are just as diverse as those of majority members. More recent research suggests that the benefits of racial diversity can be observed even before an interracial interaction occurs – independent of novel minority perspectives. One study observed improved reading comprehension in majority members even before an interracial interaction (Sommers, Warp, & Mahoney, 2008). Another demonstrated improved performance by diverse groups in the context of juror decision-making (Sommers, 2006). These studies suggest an entirely new justification for diversity-related initiatives– the benefits of diversity are experienced not just by minorities, but by all.

To discern whether there are changes in how we process information when we anticipate working with diverse others, I previously conducted a study in which participants believed they would take a logical reasoning test before and after discussing lecture material with someone either from their racial group (homogeneous interaction) or from their racial outgroup (diverse interaction) (Marotta & Sommers, 2013). The results of this study suggested a trend such that while majority students tended to outperform students of color when they anticipated a homogenous interaction, students of color significantly outperformed majority students when anticipating a heterogeneous interaction. These results contradicted the hypothesis that everyone engages in more thorough, successful processing when anticipating a diverse interaction. However, it also suggested that there may be something about racial diversity in particular that is more taxing to majority students than to students of color.

A possible explanation for this finding is that Whites in interracial interactions are concerned with appearing prejudiced and may employ color-blind strategies that limit their cognitive resources, hindering their own performance and that of their partner (Holoien & Shelton, 2012). Further, Whites may purposely avoid acknowledging race, even in contexts in which acknowledging race would improve their performance, in order to maintain an unprejudiced image (Norton, Sommers, Apfelbaum, Pura, & Ariely, 2006). It is possible that the participants in Marotta & Sommers (2013) engaged in these color-blind strategies in anticipation of their interracial interaction and that this hindered their performance. Further, since there was no actual interaction in Marotta & Sommers (2013) it is possible that had the students interacted, the performance of students of color would also have suffered. The present study replicates this study in a race neutral context in order to differentiate the general effects of diversity on cognitive processing from those that are specific to diversity with historical associations. The present study also extends the Marotta & Sommers (2013)



study by assessing whether these differences are observed when participants anticipate working in a diverse group as opposed to in a diverse dyad.

Ingroup favoritism is not limited to groups with which we have an emotional attachment (Tajfel, Billig, Bundy, & Flament, 1971). Ingroup favoritism is observed even when groups are established artificially in experimental settings based on arbitrary or even fabricated criteria. This artificial creation of groups is referred to as the minimal group paradigm and it enables researchers to observe the effects of intergroup contact between novel groups that never actually interact or have any prior history upon which to base biases (Tajfel, 1974/2010). The present study used the minimal group paradigm to create pseudo-intergroup interactions in order to observe the effects of abstract diversity on logical reasoning, creativity, and problem-solving.

Past research suggests that when diversity is managed and there is little interdependence between group members, diverse groups tend to outperform homogeneous groups (DeDreu & Weingart, 2003; Jackson & Joshi, 1999). The present study created an artificial social categorization in order to minimize the potential for group or emotional conflict while suggesting that there were psychological differences between the groups that could potentially lead to task conflict. It was hypothesized that participants anticipating a diverse interaction would outperform participants anticipating a homogeneous interaction and those who anticipated working alone (control condition) on measures of their creativity, logical reasoning, and commission of cognitive fallacies. Participants were led to believe that their performance was not dependent on that of others and thus I predicted that under these circumstances I would only observe the positive effects of diversity on performance.

Two possible causes of improved performance in diverse interactions are the fear of a negative interaction with outgroup members leading to increased attention and consideration of the materials, and attempting to compensate for an anticipated unequal distribution of effort. Also,

anticipating a diverse interaction could lead to increased arousal and a potentially negative mood. Past research suggests that increased arousal leads to more global processing, which could help in problem-solving, and that negative mood states facilitate creative problem-solving (Brunyé, Mahoney, Augstyn, & Taylor, 2009; Kaufmann & Vosberg, 2010). If participants experience gains in certain domains and not others, such as creativity, but not cognitive fallacies, it may also suggest areas of cognitive processing that may be targeted for collaborative diversity initiatives.

## **Method**

### **Design**

The present study is a between-subjects design in which group composition was manipulated (3(Group Composition: Diverse Group, Homogeneous Group, Control)). Participants were told that they would interact with only ingroup members or with ingroup and outgroup members in an online forum. There was also a third condition in which participants were told that they were in the control condition and would be working alone. All participants were randomly assigned to one of the three conditions: diverse, homogeneous, or control using the Research Randomizer (Urbaniak & Plous, 2011).

### **Participants**

One hundred and thirty two undergraduate students were recruited to participate in the full study ostensibly investigating the effects of online collaboration on task performance (60 men, 71 women, 1 unreported,  $M_{age} = 18.79$ ,  $SD = 0.93$ , 64% White). There was no basis for excluding participants as all participants were assigned to novel groups in order to manipulate whether interactions were diverse or homogeneous. All participants provided informed consent and were compensated with partial course credit.

## Materials

**Group membership manipulation.** Based on previously successful minimal group inductions, participants were told that they would be assigned to a group based on their preferences for a series of paintings (Pinter & Greenwald, 2011). Participants were asked to rate twelve paintings, six by Paul Klee and six by Wassily Kandinsky (1 = *Hate it* to 5 = *Love it*) and were told that their preferences would allow the researchers to infer their general approach to problem-solving (Tajfel, et al., 1971). The artist's signature was not visible on any of the paintings and regardless of their preferences, participants were provided with the false feedback that they preferred Paul Klee's works. Participants were told that for ease of reference, those participants who demonstrated a preference for Klee's work, and tended to approach problems deductively, would be assigned to the Blue Team; participants who demonstrated a preference for Kandinsky's work tended to approach problems inductively, and would be assigned to the Green Team. In the diverse condition participants believed they were going to interact with another student from the Blue Team and two students from the Green Team, while in the homogeneous condition they believed that they were going to interact only with students from the Blue team.

**Dependent measures.** The participants completed an assessment, described as a pre-test that was composed of two open-ended questions assessing creativity (Lamm & Trommsdorff, 1973), two questions assessing the use of the anchoring heuristic (Jacowitz & Kahneman, 1995), and eight questions assessing the ability to solve anagrams (Ventura, Shute, & Shao, 2013). Several multiple-choice questions assessed logical reasoning ability (Law School Admission Council, 2007), as well as commission of the sample-size and gambler's fallacies (Burns & Corpus, 2004; Tversky & Kahneman, 1974). The pre-test is reproduced in Appendix A.

**Creativity measures.** To assess creativity participants responded to two open-ended prompts. In the first prompt participants were asked to list positive and negative consequences that

might occur if everyone born after 1980 had an additional thumb on each hand (i.e., the Thumbs Problem); in the second prompt, they were asked to suggest ways that tourism to the United States could be increased (i.e., the Tourism Problem). These prompts were adapted from Lamm and Trommsdorff (1973) and assessed the ability to envision consequences and develop solutions, respectively.

***Anchoring heuristic.*** When estimating an unknown quantity, research suggests that we often generate an anchor, or a number we know is wrong, and then adjust our response away from it (Epley & Gilovich, 2001). For example, if we are asked to estimate the freezing point for vodka our thought process might be that: We know that vodka freezes at a lower temperature than water and that water freezes at 32°F, so vodka must freeze at a temperature lower than 32°F. The anchor in this example is 32° F and we would need to adjust downward from this anchor until we arrive at an estimate that we find acceptable. However, research suggests that there is a tendency to suggest estimates that are closer to the anchor than are realistic. Also, estimates can be influenced by whether the anchor is self-generated or experimenter provided, and whether the anchor is higher or lower than the actual number (Epley & Gilovich, 2001). In the present study, participants were asked to estimate two quantities adapted from Jacowitz and Kahneman (1995): the distance in miles from New York City to San Francisco and the number of current members in the United Nations. Participants were provided with anchors and their reliance on the anchors when making their estimates was measured.

***Anagrams.*** Adapted from Ventura, Shute, and Zhao (2013), the ability to solve anagrams was used to assess cognitive persistence. Anagrams are scrambled letter strings that participant are asked to unscramble. For example, if a participant is asked to solve the anagram “norb” they should respond with the solution “born.” Though easier anagrams can likely be solved with insight, as the letter strings are shorter, participants could systematically try different combinations of letters until

arriving at the correct answer. More correct answers indicate more persistence in trying combinations of letters.

***Logical reasoning ability.*** Logical reasoning questions taken from the Law School Admissions Prep Test (Law School Council, 2007) were posed to participants to assess their logical reasoning ability. These questions provide participants with a scenario and then requires them to indicate which outcome follows from the information they have been provided, such as which response relies on similarly flawed reasoning, or which outcome can be inferred from the information they have been provided.

***Cognitive fallacies.*** Participants were asked two questions assessing their commission of cognitive fallacies. The first question assessed insensitivity to sample size or the tendency to ignore the effect that sample size can have on the probability of a particular outcome (Tversky & Kahneman, 1974). Past research suggests that individuals tend to consider the prevalence of an outcome in a given population when predicting the likelihood that this outcome will occur in a particular sample, though smaller samples are more likely to deviate from the population parameter than larger samples. The question used in this study asked participants to indicate whether a large or small hospital is more likely to observe more than 60% of the babies born on a particular day being male, given that in a given day approximately 50% of babies born in the population are male. If a participant indicates that both hospitals are equally like to observe more than 60% of the babies being born they have neglected the effect of the sample size on the likelihood of this outcome. The correct answer is that the smaller hospital is more likely to observe more than 60% of their babies being male in a given day because if there are 15 babies born per day, only 9 of them need to be male for the percentage for that day to be 60% or higher. Alternatively, if at the larger hospital 45 babies are born per day, at least 27 of the babies need to be male for the percentage of babies born that day to be 60% or higher.

The second item measuring the commission of cognitive fallacies assessed whether when predicting a new outcome in a sequence, participants committed the gamblers or hot hand fallacies. The gambler's fallacy occurs when a streak of one outcome is observed and one predicts that an alternative outcome is now more likely than chance; the hot hand fallacy occurs when a streak is observed and one predicts that the streak is more likely to continue (Burns & Corpus, 2004). In a fair game in which these outcomes are independent what occurs previously in the sequence has no effect on the likelihood of the next outcome in the sequence. The question posed to participants was adapted from Burns & Corpus (2004). Participants were asked to imagine that they were working in a Las Vegas casino and observing the outcomes of a roulette wheel. They were asked to imagine that while all night the roulette wheel had landed on Black and Red with equal frequency, the past four spins had landed on red. The participant then indicated whether they thought that the next spin would result in Black, Red, or Black and Red with equal likelihood. Black and red are equally likely was the correct answer.

**Other measures.** Participants were also asked to respond to a series of questions measuring how much effort they exerted in completing the pre-test quickly and accurately, how much effort they predicted their team members would exert on the pre and post-test, how anxious they were in anticipation of their online interaction, and whether or not they remembered the teams to which their partners were assigned.

### **Procedure**

All participants arrived at the lab and were greeted by an experimenter who directed them to a private cubicle to complete the study's tasks. Participants were told that the study was investigating the impact of online discussion on task performance. Participants were led to believe that they would take a pre-test before discussing and reviewing the test's materials with a group of students from local universities over the internet. After their discussion, the students believed they would

take a post-test so that their performance before and after the discussion could be assessed.

Participants in the control condition were told that they would take a pre-test and that no further interaction or assessment was necessary.

Participants were then told that we were interested in assessing their general problem-solving approach prior to taking the pre-test using their evaluations of several paintings. Participants then rated the paintings and were told that they had been assigned to the Blue Team and tended to solve problems in a deductive manner. After assignment to the Blue Team, participants were asked to exchange some information with their partners. This information exchange was our manipulation of group composition and how participants discovered that they were interacting entirely with other Blue Team members (homogeneous interaction) or with one Blue Team and two Green Team members (diverse interaction). Next, participants were given exactly fifteen minutes to complete the test to ensure that the task was sufficiently difficult and to avoid ceiling effects. Participants were then be debriefed, compensated, and dismissed.

## Results

A linear regression was conducted to determine if individual differences between participants could predict any of my dependent variables. Results suggest that there was no statistically significant relationship between age, gender, race, or English proficiency and any of the dependent measures,  $p$ 's  $> 0.05$ . Individual differences are not discussed further, or taken into consideration in the following analyses.

### Manipulation Checks

**Group composition.** To assess whether participants believed they would be interacting with students from other universities via an internet chatroom, all participants were asked whether they believed the interaction would occur. Of the participants that were told they would be

interacting with other students, 64 % of the sample believed that the interaction would occur. Of the students that were told they were in the control condition and would be interacting alone 100% believed that this was case. The following analyses were conducted with the responses from participants that believed they would be interacting with other students over the internet.<sup>1</sup>

**Team membership.** Each participant completed an image rating task and was ostensibly assigned to a team based on their preferences for either Wassily Kandinsky or Paul Klee's artwork. The results of a T-Test suggest that in general, participants preferred Wassily Kandinsky's work ( $M = 3.16, SD = 0.57$ ) to that of Paul Klee ( $M = 2.89, SD = 0.5766$ ),  $t(128) = 5.165, p < 0.001$ . However, there was no statistically significant effect of condition on liking of Kandinsky's work,  $F(2, 126) = 0.30, p = 0.75, ns$ , or on liking of Klee's work,  $F(2, 126) = 0.05, p = 0.95, ns$ .

## Dependent Measures

I predicted that participants anticipating a diverse interaction would outperform participants anticipating a homogeneous interaction on measures of creativity, logical reasoning, and reliance on cognitive heuristics. To assess this hypothesis, one-way ANOVAs were conducted with Group Composition as a fixed factor and each of the dependent measures as a dependent factor in turn (see *Table 1*). For measures with count data,  $\chi^2$  analyses were conducted (see *Table 2*).

**Creativity measure.** The responses that each participant provided to either prompt were counted to indicate the ease with which participants could suggest answers. Further, responses to the Thumbs Problem were separated into positive and negative consequences. Responses were then coded by independent, naive, blind, raters on the probability that these consequences would occur (Cronbach's  $\alpha_{\text{positive}} = 0.53$ ; Cronbach's  $\alpha_{\text{negative}} = 0.57$ ) and how important those consequences

---

<sup>1</sup> Including the participants that did not believe the manipulation in the analyses did not change the results substantially.



would be for society (Cronbach's  $\alpha_{\text{positive}} = 0.70$ ; Cronbach's  $\alpha_{\text{negative}} = 0.70$ ). Responses to the Tourism Problem were coded by independent, naive, blind raters for their effectiveness (Cronbach's  $\alpha = 0.65$ ) and feasibility (Cronbach's  $\alpha = 0.77$ ).

***Thumbs problem- Positive consequences.*** Results suggest that participants in the control condition ( $M = 2.49, SD = 1.12$ ) suggested more positive consequences than those in the homogeneous condition ( $M = 2.37, SD = 1.03$ ) and those in the diverse condition ( $M = 2.10, SD = 0.80$ ),  $F(2, 94) = 1.25, p = 0.29, ns$ , but this difference was not statistically significant. Participants in the control condition ( $M = 4.81, SD = 0.93$ ) suggested consequences that were scored as more probable than those participants in the diverse ( $M = 4.80, SD = 0.86$ ) and homogeneous ( $M = 4.66, SD = 1.08$ ) conditions, but this difference was not statistically significant,  $F(2, 94) = 0.24, p = 0.79, ns$ . Participants in the control condition ( $M = 3.91, SD = 1.09$ ) provided positive responses that were scored as more important for society than those provided by participants in the diverse ( $M = 3.57, SD = 0.85$ ) and homogeneous conditions ( $M = 3.48, SD = 1.07$ ), but these differences were not statistically significant,  $F(2, 94) = 1.71, p = 0.19, ns$ .

***Thumbs problem- Negative consequences.*** Results suggest that participants in the control condition ( $M = 2.64, SD = 1.10$ ) suggested more negative consequences than those in the homogeneous condition ( $M = 2.27, SD = 0.91$ ) and those in the diverse condition ( $M = 2.23, SD = 0.94$ ),  $F(2, 93) = 1.74, p = 0.18, ns$ , but this trend was not statistically significant. Participants in the diverse condition ( $M = 3.97, SD = 0.82$ ) suggested consequences that were scored as more probable than those suggested by participants in the homogeneous ( $M = 3.73, SD = 0.85$ ) and control ( $M = 3.43, SD = 0.99$ ) conditions, but this difference was not statistically significant,  $F(2, 94) = 1.59, p = 0.21, ns$ . Participants in the diverse condition ( $M = 4.17, SD = 0.69$ ) provided negative responses that were scored as more important for society than those provided by participants in the control ( $M$

= 3.88,  $SD = 0.83$ ) and homogeneous conditions ( $M = 3.79$ ,  $SD = 0.86$ ), but these differences were not statistically significant,  $F(2, 94) = 1.94$ ,  $p = 0.15$ , *ns*.

***Tourism problem.*** The second open-ended prompt asked participants to suggest ways to increase tourism to the United States. Participants in the diverse condition suggested more solutions ( $M = 2.87$ ,  $SD = 0.94$ ) than those in the control ( $M = 2.86$ ,  $SD = 1.40$ ) and homogeneous conditions ( $M = 2.70$ ,  $SD = 1.18$ ), but this difference was not statistically significant,  $F(2, 94) = 0.19$ ,  $p = 0.83$ , *ns*. Participants in the diverse condition provided responses that were perceived as more feasible ( $M = 4.64$ ,  $SD = 0.97$ ) than participants in the control ( $M = 4.53$ ,  $SD = 1.15$ ) and homogeneous conditions ( $M = 4.09$ ,  $SD = 1.10$ ), but this difference was not statistically significant,  $F(2, 93) = 2.24$ ,  $p = 0.11$ , *ns*. However, a statistically significant effect of group composition on the effectiveness of tourism solutions was observed in which the solutions proposed by participants in the diverse condition were perceived as more effective ( $M = 4.63$ ,  $SD = 0.83$ ) than those suggested by participants in the control ( $M = 4.26$ ,  $SD = 0.86$ ) and homogeneous conditions ( $M = 4.10$ ,  $SD = 0.91$ ),  $F(2, 93) = 3.02$ ,  $p = 0.05$ . The results of a post-hoc Tukey statistical test indicate that participants in the diverse condition performed statistically significantly better than those in the homogeneous condition,  $p = 0.05$ , but there was no difference in performance between the diverse and control conditions or between the homogeneous and control conditions,  $p$ 's  $> 0.20$ .

While on some measures, participants anticipating a diverse interaction tended to outperform those anticipating a homogeneous interaction or no interaction, these findings were not statistically significant or consistent in most cases. These results do not support my hypothesis that anticipating a diverse interaction will lead to improved performance on creativity measures. However, on no measure of creativity did participants anticipating a homogeneous interaction

outperform both participants anticipating a diverse interaction and those not anticipating an interaction.

**Anchoring heuristic.** To assess the reliance on the anchoring heuristic Epley and Gilovich's (2001) procedure was used to calculate the average anchoring bias score for each question. First, imprecise answers, such as "less than 1600" were removed from the data set. Then, the absolute value of the anchor less the participant's estimate was calculated and these scores served as the measure of anchoring bias. Lower numbers therefore indicate less deviation from the anchor and higher numbers indicate more deviation from the anchor and less reliance on the anchoring heuristic.

For the first question, which asked participants to estimate the distance from San Francisco to New York in miles, participants in the control condition demonstrated less reliance on the anchor ( $M = 3471.19$ ,  $SD = 5835.65$ ) than participants in the diverse ( $M = 2597.75$ ,  $SD = 926.78$ ) and homogeneous conditions ( $M = 2096.82$ ,  $SD = 1063.33$ ), but this difference was not statistically significant,  $F(2, 90) = 1.11$ ,  $p = 0.33$ , *ns*. For the second question, which asked participants to estimate the current number of United Nations members, the opposite trend was observed with participants in the homogeneous condition ( $M = 61.62$ ,  $SD = 67.61$ ) demonstrating less reliance on the anchor than participants in the diverse ( $M = 59.10$ ,  $SD = 65.30$ ) and control ( $M = 55.57$ ,  $SD = 60.61$ ) conditions, but this difference was not statistically significant,  $F(2, 93) = 0.07$ ,  $p = 0.93$ , *ns*. The results of this measure do not support the hypothesis that participants anticipating a diverse interaction will outperform those anticipating a homogeneous interaction or working alone on avoiding the anchoring heuristic.

**Anagrams.** The Anagram Accuracy score was computed as the number of correctly solved anagrams divided by the number of anagrams attempted and multiplied by 100. Participants in the

homogeneous condition ( $M = 89.24, SD = 18.67$ ) outperformed participants in the diverse ( $M = 82.66, SD = 18.96$ ) and control conditions ( $M = 82.60, SD = 19.03$ ), but this difference was not statistically significant,  $F(2, 92) = 1.26, p = 0.29, ns$ . The results of this measure do not support the hypothesis that participants anticipating a diverse interaction will outperform those anticipating a homogeneous interaction or working alone on tasks of persistence.

**Logical reasoning.** The logical reasoning ability of participants was assessed using a similar method to that of the anagrams. The number of correctly answered logical reasoning questions was divided by the number of questions attempted and multiplied by 100 to serve as the Logical Reasoning Accuracy Score (Law School Council, 2007). Results suggest a statistically significant effect of group composition condition on logical reasoning accuracy such that participants in the diverse condition ( $M = 93.33, SD = 15.99$ ) outperformed participants in the homogeneous condition ( $M = 84.17, SD = 19.22$ ) and the control condition ( $M = 81.76, SD = 20.96$ ),  $F(2, 94) = 3.31, p = 0.04$ . The results of a post hoc Tukey statistical test suggest that participants in the diverse condition performed significantly better than those in the control condition,  $p = 0.04$ , but performed similarly to those in the homogeneous condition,  $p = 0.15$ . Additionally, there was no difference in performance on this measure between participants in the homogeneous and control conditions,  $p = 0.86$ . This lends partial support to my hypothesis that participants anticipating a diverse interaction will outperform those anticipating a homogeneous interaction or working alone on tasks of logical reasoning.

**Cognitive fallacies.** Responses were coded to indicate if a participant had committed the insensitivity to sample size, gambler's, or hot-hand fallacies. These coded responses were then submitted to a  $\chi^2$  analysis to determine if there was a difference in commission of each fallacy by

group composition condition<sup>2</sup>. The results of this analysis suggest that across conditions, participants were equally likely to commit the insensitivity to sample size fallacy  $\chi^2(2, N = 97) = 0.90, p = 0.93, ns$ . Also, across conditions participants were equally likely to commit the gamblers and hot hand fallacies,  $\chi^2(4, N = 96) = 2.45, p = 0.66, ns$ . These results do not support the hypothesis that participants anticipating a diverse interaction will be less likely to rely on cognitive heuristics.

**Possible mediators.** To begin determining which factors may be influencing performance based on whether the participant anticipates working alone or with a homogeneous or diverse group, participants were asked to respond to several questions assessing their feelings about their upcoming interaction, if applicable. Participants in the experimental conditions were asked to report how anxious they felt about the upcoming interaction. Participants in the homogeneous condition ( $M = 2.97, SD = 1.59$ ) reported feeling more anxious about the upcoming interaction than participants in the diverse condition, ( $M = 2.43, SD = 1.57$ ) but this was not a statistically significant difference  $F(1, 55) = 1.64, p = 0.21, ns$ . These participants were also asked to report how optimistic they felt about the upcoming interaction. Participants in the diverse condition reported more optimism about interacting with the group ( $M = 5.00, SD = 0.77$ ) than participants in the homogeneous condition, ( $M = 4.97, SD = 0.78$ ) but this difference was not statistically significant  $F(1, 55) = 0.03, p = 0.87, ns$ .

To determine if there were differences in how much participants felt their group members would contribute to the interaction, participants were asked to report how much effort they felt each of their group members would put into the post-test. To compute a score for Group Effort, the ratings for each group member were averaged (Cronbach's  $\alpha = 0.95$ ). Results suggest that

---

<sup>2</sup> All expected counts were not greater than or equal to 5; however, collapsing conditions to increase the expected cell counts did not affect the results of the  $\chi^2$  analysis.

participants in the homogeneous condition ( $M = 5.02, SD = 0.81$ ) predicted that their groups would put more effort into the post-test than participants in the diverse condition, ( $M = 4.95, SD = 0.91$ ) but this difference was not statistically significant  $F(1, 55) = 0.12, p = 0.73, ns$ . In addition to estimating the effort of their group members, each participant reported their own effort on the pre-test and projected effort on the post-test (when applicable). Participants in the diverse condition ( $M = 5.23, SD = 0.77$ ) reported putting more effort into the test than participants in the homogeneous ( $M = 5.07, SD = 0.83$ ) or control conditions; ( $M = 5.06, SD = 1.09$ ); however, these differences were not statistically significant,  $F(2, 85) = 0.31, p = 0.73, ns$ .

**Post study.** After data collection for the present study was complete, additional data were collected to determine if aspects of the experimental design may have influenced results in unanticipated ways. Ninety-six participants (37 men, 59 women,  $M_{age} = 35.01, SD = 12.69, 75\%$  White) were recruited from Amazon's Mechanical Turk work force to complete an abbreviated version of the main study. In this post study, participants were randomly assigned to either the Blue or Green team, ostensibly based on their painting preferences. Participants were then asked to respond to a series of questions assessing how similar they felt to other people on their team, their preferences for the Blue versus Green teams, and how much they believed that their problem-solving approach could be determined by the painting preferences. Participants were then debriefed and compensated \$0.25.

Data from 12 participants were removed from the following analyses based on failure to correctly indicate the team they had been assigned to and their own admission of not attending to the study materials. Analyses are based on the data from the remaining 84 participants (31 men, 53 women,  $M_{age} = 35.33, SD = 13.03, 75\%$  White). The results of an independent samples T-Test suggest that participants assigned to the Blue Team ( $M = 5.33, SD = 1.00$ ) felt more similar to other

individuals assigned to the Blue Team than did participants assigned to the Green Team ( $M = 3.72$ ,  $SD = 1.38$ ),  $t(82) = 6.21$ ,  $p < 0.001$ . Similarly, participants assigned to the Green Team ( $M = 4.85$ ,  $SD = 1.07$ ) felt more similar to other individuals assigned to the Green Team than did participants assigned to the Blue Team ( $M = 2.98$ ,  $SD = 1.36$ ),  $t(82) = -6.94$ ,  $p < 0.001$ . This suggests that participants did experience some affiliation with their assigned teams. Results also suggest that participants on the Blue Team ( $M = 5.07$ ,  $SD = 1.10$ ) demonstrated a stronger preference for being a deductive thinker than participants assigned to the Green Team ( $M = 4.51$ ,  $SD = 0.85$ ),  $t(82) = 2.55$ ,  $p = 0.01$ . However, as the participants that were told they were deductive thinkers (Blue Team) demonstrated this preference this likely did not influence my results. Participants assigned to Green team ( $M = 4.56$ ,  $SD = 1.02$ ) tended to demonstrate a stronger preference for being an inductive thinker than participants assigned to the Blue Team ( $M = 4.24$ ,  $SD = 1.13$ ), but this difference was not statistically significant  $t(82) = -1.35$ ,  $p = 0.18$ , *ns*. There was also no statistically significant difference between the participants assigned to the Blue Team ( $M = 3.87$ ,  $SD = 1.49$ ) and those assigned to the Green Team ( $M = 3.46$ ,  $SD = 1.48$ ) in terms of how much they believed their problem-solving approach could be determined by their painting preferences,  $t(82) = 1.25$ ,  $p = 0.22$ . Based on these results, it is unlikely that assigning all participants to the Blue (deductive thinkers) team influenced the results of the main study.

### Discussion

The results of this study suggest that there are several domains in which anticipating an interaction with a diverse group may improve our cognitive performance, but this finding was neither statistically significant nor consistent across all measures. On most measures of creativity, and all the measures of cognitive persistence and reliance on cognitive heuristics, performance was comparable across conditions and no statistically significant effects were observed. However, on one measure of creativity, the effectiveness of proposed solutions to the Tourism Problem,

participants in the diverse condition did perform statistically significant better than participants in the homogeneous condition, as predicted; although the performance of participants in the diverse condition was comparable to that of those in the control condition. Additionally, on the measure of logical reasoning participants in the diverse condition performed statistically significantly better than those in the control condition, though their performance was comparable to that of participants in the homogeneous conditions. These results lend partial support to the hypothesis that prior to an interaction participants anticipating a diverse interaction will demonstrate improved task performance when diversity is manipulated in an abstract way.

The two significant findings that participants in the diverse condition developed more effective solutions than those in the homogeneous condition, and that participants in the diverse condition demonstrated better logical reasoning ability than participants in the control condition, suggest that there may be circumstances under which merely anticipating a diverse interaction can improve how we approach problems. This is somewhat consistent with the finding in Marotta and Sommers (2013) which suggested that students of color performed better on a logical reasoning task when anticipating working with White as opposed to Black students. The present study extends this finding by determining if these gains can be demonstrated in a scenario in which diversity is not manipulated by race or another social category, but by an arbitrarily assigned team membership: problem solving approach.

Though there were two statistically significant findings of note, most of the predicted differences among conditions were not observed. This may be partially explained by some of the differences between the study designs used in past research and that used in the present study. For example, in the Sommers, et al. (2008) study, improvements in performance were observed in anticipation of interacting in a diverse group, but the participants were completing the tasks in the same room as their ostensible partners. In the present study, all participants completed the study in a



private cubicle and only directly interacted with the experimenter. The use of confederates in past research may have added realism to the scenario that was absent from the present study. Though online interactions are frequent and increasingly commonplace in the workplace and higher education it may be the case that this anticipated interaction was not realistic enough to induce changes in performance. Future studies could more closely replicate the study by Sommers and colleagues (2008) by employing confederates. However, as originally conceived when using the minimal group paradigm the participants should not interact so that preferences or biases cannot be made based on differences other than those created by the paradigm and a different operationalization of diversity might be needed (Tajfel, 1974/2010). Another option could be to continue with the use of an online interaction but to use avatars to manipulate diversity, which would add a visible representation of the ostensible partners and possibly make the scenario more believable.

In addition to the difference in mundane realism between the present study and past research, in past research improvements in performance were only observed when the topic that participants would be discussing was related to the dimension of diversity that they were experiencing (i.e. needing to discuss race in a racially diverse group) (Sommers, et al., 2008). In the present study, the material the participants were reviewing was devoid of controversial issues and generally independent of the diversity dimension they were experiencing (deductive versus inductive thinker). It may be the case that a stronger social identity and the need to complete a task relevant to that identity are necessary to observe differences in performance. In support of this possibility, it is noteworthy that significant findings were observed on the measures upon which being a deductive thinker would be most relevant, developing effective solutions and reasoning logically. Future studies could further assess the degree to which anticipating diversity based on psychological

tendencies can influence our behavior in subsequent interactions, and our performance on related tasks.

Another concern in this study was the frequency with which participants in the control condition outperformed those in the experimental conditions. On most of the measures of creativity and the anchoring heuristic the participants in the control condition tended to outperform participants in the experimental conditions. Though these trends were not statistically significant, it may suggest that group work in general is taxing for some students, regardless of the composition of the group. Further investigation is warranted to determine the elements of a group interaction, in general, that can harm the performance of students. Additionally, on the measure of cognitive persistence participants in the homogeneous condition tended to outperform those in the control and diverse conditions. Though this trend was not statistically significant, it may be explained by the higher rates of turnover in diverse groups

While the results of the present study are inconclusive, further investigation of the influence of diversity is necessary to better understand the social and political changes the nation is experiencing. For example, affirmative action policies were implemented in the 1960s to correct for historical injustices and to provide minorities and other protected classes with opportunities in hiring and higher education (Garrison & Modigliani, 1994). However, these policies have always been perceived as temporary remedies to a societal condition. Recent Supreme Court decisions that sustain these policies predicted that by 2028 considering factors like race in college admissions would no longer be necessary (*Grutter v. Bollinger*, 2003). The changing demographic of the United States, the election of a biracial president, and media propagation of the United States as a nation beyond racial concerns may further suggest that social policies like affirmative action are no longer needed (Taylor, 2014). However, the American Psychological Association's Presidential Task Force on Educational Disparities reports that there are still gross ethnic and racial disparities in both K-12

and higher education. Thus research that can disambiguate the benefits and costs of diversity in these domains is necessary and important.

While this study begins to examine the circumstances under which one can anticipate positive versus negative outcomes as a result of diverse interactions, there remains a gap in the literature that addresses the mechanism that is driving the observed changes in performance. The present study began this investigation by asking participants to report their anxiety and optimism concerning the upcoming interaction and the effort they predicted their partners would put forth. The observed trend suggested that participants anticipating a diverse interaction were both more anxious and more optimistic about the interaction than those in the homogeneous condition, though this was not a statistically significant difference. These emotional states might be affecting performance on cognitive tasks as demonstrated in previous studies (Brunyé, Mahoney, Augstyn, & Taylor, 2009; Kaufmann & Vosberg, 2010). However, future studies could further investigate these issues while inducing mood states or manipulating arousal to determine if this affects performance on cognitive tasks differently depending on whether one is anticipating working with a diverse or homogeneous group.

Additionally, participants in the diverse condition tended to report that they would put forth more effort than those in the homogeneous condition, though this was not a statistically significant difference. This is paralleled by the pattern suggesting that participants in the homogeneous condition predicted that their partners would contribute more effort than those in the diverse condition, though this too was not a statistically significant difference. It is not clear whether or not the increased effort in the diverse condition is a defense against the anticipated lesser effort of one's group, or is the result of another mechanism. Future research could manipulate the effort contributed by group members in addition to manipulating the group's composition to parse out this issue further. It may be the case that participants are working harder in diverse groups because

they do not expect their group mates to contribute; however, if their group mates are actually contributing the group's overall performance could improve.

The present study begins a program of research that could provide evidence that everyone has the potential to benefit from diversity and could suggest ways in which the negative consequences of working in diverse groups can be mitigated. This provides a compelling reason to continue programs that encourage diversity in education and hiring, despite the strides toward equality and correcting historical injustices that have already occurred. As our nation further diversifies, working with diverse others will become increasingly unavoidable. It is critical that we find ways to exploit the benefits associated with these changes and minimize the taxing and emotionally distressing consequences in order to protect and further the progress that has been made toward a more egalitarian nation.

## References

- American Psychological Association, Presidential Task Force on Educational Disparities. (2012). *Ethnic and racial disparities in education: Psychology's contributions to understanding and reducing disparities*. Retrieved from <http://www.apa.org/ed/resources/racial-disparities.aspx>
- Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. (2004). Effects of racial diversity on complex thinking in college students. *Psychological Science, 15*, 507-510.
- Brunyé, T. T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). Emotional state and local versus global spatial memory. *Acta Psychologica, 130*, 138–146.
- Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: “Gambler’s fallacy” versus “hot hand”. *Psychonomic Bulletin & Review, 11*(1), 179-184.
- Cunningham, G.B. (2009). The moderating effect of diversity strategy on the relationship between racial diversity and organizational performance. *Journal of Applied Social Psychology, 39*(6), 1445-1460.
- DeDreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology, 88*(4), 741–749.  
doi:10.1037/0021-9010.88.4.741
- Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science, 12*(5), 391-396.
- Garrison, W. A. & Modigliani, A. (1994). The changing culture of affirmative action. *Equal Employment Opportunity: Labor Market Discrimination and Public Policy, 373-394*.
- Greenwald, A. G., Pickrell, J. E., & Farnham, S. D. (2002). Implicit partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology, 83*(2), 367-379.
- Grutter v. Bollinger, 539 U.S. 306 (2003).

- Harvey, S. (2013). A different perspective: The multiple effects of deep level diversity on group creativity. *Journal of Experimental Social Psychology, 49*, 822–832.
- Holoien, D. S., & Shelton, J. N. (2012). You deplete me: The cognitive costs of colorblindness on ethnic minorities. *Journal of Experimental Social Psychology, 48*(2), 562-565. doi: 10.1016/j.jesp.2011.09.010
- Jackson, S. E., & Joshi, A. (2011). Work team diversity. *APA Handbook of Industrial and Organizational Psychology, Vol 1: Building and Developing the Organization*, 651–686. doi:10.1037/12169-020
- Jacowitz, K.E. & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*(11) 1161-1166.
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly, 44*(4), 741–763.
- Kaufmann, G., & Vosburg, S. K. (1997). “Paradoxical” mood effects on creative problem-solving. *Cognition and Emotion, 11*(2), 151–170. doi: 10.1080/026999397379971
- Kerr, N. & Tinsdale, R.S. (2004). Group Performance and Decision Making. *Annual Review of Psychology, 55*, 623–655.
- Konan, P. N., Chatard, A., Selimbegović, L., & Mugny, G. (2010). Cultural diversity in the classroom and its effects on academic performance: A cross-national perspective. *Social Psychology, 41*(4), 230-237.
- Lamm, H. & Trommsdorff, G. (1973). Group versus individual performance on tasks requiring ideation proficiency (brainstorming): A review. *European Journal of Social Psychology, 3*(4), 361-388.

- Law School Admission Council (2007). *The official LSAT prep-test*. Retrieved October 3, 2012 from <http://www.lsac.org/jd/pdfs/sampleptjune.pdf>.
- McLeod, P.L., Lobel, S.A., & Cox, Jr., T.H. (1996). Ethnic diversity and creativity in small groups. *Small Group Research*, 27(2), 248-264.
- Mannix, E. & Neale, M. A. (2005). What Differences Make a Difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest*, 6(2), 31–55. doi: 10.2307/40062326
- Marotta, S. & Sommers, S. (2013). *Does diversity make us “smarter?” Racial diversity and logical reasoning*. Unpublished manuscript, Department of Psychology, Tufts University, Medford, United States of America.
- Norton, M., & Sommers, S. (2011). Whites see racism as a zero-sum game that they are now losing. *Perspectives on Psychological Science*, 6(3), 215–218.
- Norton, M.I., Sommers, S.R., Apfelbaum, E.P., Pura, N., & Ariely, D. (2006). Color blindness and interracial interaction: Playing the political correctness game. *Psychological Science*, 17(11), 949-953.
- Outten, H. R., Schmitt, M. T., Miller, D. A., & Garcia, A. L. (2012). Feeling threatened about the future: White’s emotional reactions to anticipated ethnic demographic changes. *Personality and Social Psychology Bulletin*, 38(1), 14–25. doi: 10.1177/0146167211418531
- Phillips, K. W., Liljenquist, K. A., & Neale, M. A. (2009). Is the pain worth the gain? The advantages and liabilities of agreeing with socially distinct newcomers. *Personality and Social Psychology Bulletin*, 35(3), 336–350. doi: 10.1177/0146167208328062
- Pinter, B. & Greenwald, A.G. (2011). A comparison of minimal group induction procedures. *Group Processes Intergroup Relations*, 14(1), 81–98.

- Plaut, V. C., Thomas, K. M., & Goren, M. J. (2009). Is multiculturalism or color blindness better for minorities? *Psychological Science*, 20(4), 444–446.
- Sommers, S. R. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology*, 90(4), 597-612.
- Sommers, S. R., Warp, L. S., & Mahoney, C. C. (2008). Cognitive effects of racial diversity: White individuals' information processing in heterogeneous groups. *Journal of Experimental Social Psychology*, 44, 1129-1136.
- Tajfel, H. (2010) Social identity and intergroup behavior. In T. Postmes & N.R. Branscombe (Eds.), *Rediscovering Social Identity: Key Readings in Social Psychology* (pp.77-94). New York, NY: Psychology Press. (Reprinted from Tajfel, H. (1974). *Social Science Information/Information sur les Sciences Sociales*, 13(2), pp. 65-93).
- Tajfel, H., Billig, M. G., Bundy, R. P. and Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2), 149–178.
- Taylor, P. (April, 2014). “The Next America.” Pew Research Center, Washington, D.C.  
<http://www.pewresearch.org/next-america/#Two-Dramas-in-Slow-Motion>, accessed on May 4, 2014.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1130.
- Urbaniak, G. C., & Plous, S. (2011). Research Randomizer (Version 3.0) [Computer software].
- van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, 58, 515–541. doi:10.1146/annurev.psych.58.110405.085546
- Ventura, M., Shute, V., & Zhao, W. (2013). Anagram – riddle task [Database record]. Retrieved from PsycTests. doi: 10.1037/t18777-000



Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*, 20, 77–140.

Table 1

Means and ANOVA Statistics for Dependent Measures

| MEASURE               | QUESTION/ITEM  | GROUP COMPOSITION |           |          |           |             |           | <i>df</i> | <i>F</i> | <i>p</i> |
|-----------------------|----------------|-------------------|-----------|----------|-----------|-------------|-----------|-----------|----------|----------|
|                       |                | Control           |           | Diverse  |           | Homogeneous |           |           |          |          |
|                       |                | <i>M</i>          | <i>SD</i> | <i>M</i> | <i>SD</i> | <b>M</b>    | <b>SD</b> |           |          |          |
| Anagrams              | Accuracy Score | 82.60             | 19.03     | 82.66    | 18.96     | 89.24       | 18.67     | (2,92)    | 1.26     | 0.29     |
| Anchoring Heuristic   | Mileage        | 3471.19           | 5835.65   | 2597.75  | 926.75    | 2096.82     | 106.33    | (2,90)    | 1.11     | 0.33     |
|                       | UN members     | 55.57             | 60.61     | 59.10    | 65.30     | 61.62       | 67.61     | (2,93)    | 0.07     | 0.93     |
| Logical Reasoning     | Accuracy Score | 81.76             | 20.96     | 93.33    | 15.99     | 84.17       | 19.22     | (2,94)    | 3.31     | 0.04     |
| Thumbs Problem        | Count          | 2.49              | 1.12      | 2.10     | 0.80      | 2.37        | 1.03      | (2,94)    | 1.25     | 0.29     |
| Positive Consequences | Probability    | 4.81              | 0.93      | 4.80     | 0.86      | 4.66        | 1.08      | (2,94)    | 0.24     | 0.79     |
|                       | Importance     | 3.91              | 1.09      | 3.57     | 0.85      | 3.48        | 1.07      | (2,94)    | 1.71     | 0.19     |
| Negative Consequences | Count          | 2.64              | 1.10      | 2.23     | 0.94      | 2.27        | 0.91      | (2,93)    | 1.74     | 0.18     |
|                       | Probability    | 3.43              | 0.99      | 3.97     | 0.82      | 3.73        | 0.85      | (2,94)    | 1.59     | 0.21     |
|                       | Importance     | 3.88              | 0.83      | 4.17     | 0.69      | 3.79        | 0.86      | (2,94)    | 1.94     | 0.15     |
| Tourism Problem       | Count          | 2.86              | 1.40      | 2.87     | 0.94      | 2.70        | 1.18      | (2,93)    | 0.19     | 0.83     |
|                       | Feasibility    | 4.53              | 1.15      | 4.64     | 0.97      | 4.09        | 1.10      | (2,93)    | 2.24     | 0.11     |
|                       | Effectiveness  | 4.26              | 0.86      | 4.63     | 0.83      | 4.10        | 0.91      | (2,93)    | 3.02     | 0.05     |

*Note.* Means and standard deviation of participant responses, or rater scores, to each question.

Table 2

Counts and  $\chi^2$  Statistics for Dependent Measures

| Measure            | Item      | Group Composition |                |         |                |             |                | <i>N</i> | <i>df</i> | $\chi^2$ | <i>p</i> |
|--------------------|-----------|-------------------|----------------|---------|----------------|-------------|----------------|----------|-----------|----------|----------|
|                    |           | Control           |                | Diverse |                | Homogeneous |                |          |           |          |          |
|                    |           | Count             | Expected Count | Count   | Expected Count | Count       | Expected Count |          |           |          |          |
| Gambling Fallacies |           |                   |                |         |                |             |                | 96       | 4         | 2.44     | 0.66     |
|                    | Correct   | 34                | 33.1           | 25      | 26.9           | 27          | 26             |          |           |          |          |
|                    | Gambler's | 2                 | 3.1            | 4       | 2.5            | 2           | 2.4            |          |           |          |          |
|                    | Hot-Hand  | 1                 | 0.8            | 1       | 0.6            | 0.0         | 0.6            |          |           |          |          |
| Sample Size        |           |                   |                |         |                |             |                | 97       | 2         | 0.90     | 0.93     |
|                    | Correct   | 19                | 17.9           | 15      | 14.5           | 13          | 14.5           |          |           |          |          |
|                    | Incorrect | 3                 | 3.1            | 3       | 2.5            | 2           | 2.5            |          |           |          |          |
|                    | Fallacy   | 15                | 16             | 12      | 13             | 15          | 13             |          |           |          |          |

*Note.* Counts indicate the number of participants that provided the corresponding response to the question on the pre-test. Correct counts indicate that the participant avoided the cognitive fallacy.

## Appendix A

**BROWNIE BATTER  
PRE-TEST**

You have 15 minutes to respond to multiple choice questions and open-ended questions. The experimenter will notify you when you have five minutes remaining and when the pre-testing session is concluded. Please respond as quickly, accurately, and thoughtfully as possible. Your score on the pre-test will be compared to your score on the post-test after you have had the opportunity to discuss the material with your partners in the online chat room.



4. Please solve the following anagrams. Anagrams are a string of letters that can be reordered to make a word. For example, YBO would be an anagram for BOY. Your task is to rearrange the letters into a word and type in your guess. Some words will be harder than others but try to do your best to make a word out of the anagram. If you do not know an answer you may guess, or leave the textbox blank.
- a) lalb \_\_\_\_\_
  - b) yaaw \_\_\_\_\_
  - c) nidk \_\_\_\_\_
  - d) nomo \_\_\_\_\_
  - e) bluc \_\_\_\_\_
  - f) naric \_\_\_\_\_
  - g) spey \_\_\_\_\_
  - h) styrt \_\_\_\_\_
5. A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?
- a) The larger hospital
  - b) The smaller hospital
  - c) About the same (that is within 5% of each other)
6. Over summer break, you find a job in Vegas where you will be an undercover security guard dressed as Elvis at some casino. After the management hands you the brightly sequined jumpsuit and fake sideburns, he tells you that your job will be to watch the roulette table. He gives you a brief description of roulette and tells you that there are basically numbers and colors (red or black) you can bet on. After your first day on the job, you are told to watch the security tape of that night's business. You know that, throughout the night, 50 out of 100 of the times the wheel stopped on red. About halfway through the tape, your boss comes in to give you a cup of

coffee and keep you company. While you are watching the tape, the wheel lands on red the next 4 times in a row. What do you predict the wheel will land on next, black or red?

- a) Black
- b) Red
- c) Black or Red are equally likely

7. What is the distance from San Francisco to New York (in miles)?

Hint: It is less than 6,000 miles.

Estimated Distance: \_\_\_\_\_

8. What is the number of United Nations members?

Hint: There are more than 14 members.

Estimated Number of Members: \_\_\_\_\_

9. All Labrador retrievers bark a great deal. All Saint Bernards bark infrequently. Each of Rosa's dogs is a cross between a Labrador retriever and a Saint Bernard. Therefore, Rosa's dogs are moderate barkers.

Which one of the following uses flawed reasoning that most closely resembles the flawed reasoning used in the argument above?

- a) All students who study diligently make good grades. But some students who do not study diligently also make good grades. Jane studies somewhat diligently. Therefore, Jane makes somewhat good grades.
- b) All type A chemicals are extremely toxic to human beings. All type B chemicals are nontoxic to human beings. This household cleaner is a mixture of a type A chemical and a type B chemical. Therefore, this household cleaner is moderately toxic.
- c) All students at Hanson School live in Green County. All students at Edwards School live in Winn County. Members of the Perry family attend both Hanson and Edwards. Therefore, some members of the Perry family live in Green County and some live in Winn County.

- d) All transcriptionists know shorthand. All engineers know calculus. Bob has worked both as a transcriptionist and as an engineer. Therefore, Bob knows both shorthand and calculus.
  - e) All of Kenisha's dresses are very well made. All of Connie's dresses are very badly made. Half of the dresses in this closet are very well made, and half of them are very badly made. Therefore, half of the dresses in this closet are Kenisha's and half of them are Connie's.
10. During the construction of the Quebec Bridge in 1907, the bridge's designer, Theodore Cooper, received word that the suspended span being built out from the bridge's cantilever was deflecting downward by a fraction of an inch (2.54 centimeters). Before he could telegraph to freeze the project, the whole cantilever arm broke off and plunged, along with seven dozen workers, into the St. Lawrence River. It was the worst bridge construction disaster in history. As a direct result of the inquiry that followed, the engineering "rules of thumb" by which thousands of bridges had been built around the world went down with the Quebec Bridge. Twentieth-century bridge engineers would thereafter depend on far more rigorous applications of mathematical analysis.

Which one of the following statements can be properly inferred from the passage?

- a) Bridges built before about 1907 were built without thorough mathematical analysis and, therefore, were unsafe for the public to use.
- b) Cooper's absence from the Quebec Bridge construction site resulted in the breaking off of the cantilever.
- c) Nineteenth-century bridge engineers relied on their rules of thumb because analytical methods were inadequate to solve their design problems.
- d) Only a more rigorous application of mathematical analysis to the design of the Quebec Bridge could have prevented its collapse.
- e) Prior to 1907 the mathematical analysis incorporated in engineering rules of thumb was insufficient to completely assure the safety of bridges under construction.



11. Ethicist: The most advanced kind of moral motivation is based solely on abstract principles. This form of motivation is in contrast with calculated self-interest or the desire to adhere to societal norms and conventions.

The actions of which one of the following individuals exhibit the most advanced kind of moral motivation, as described by the ethicist?

- a) Bobby contributed money to a local charity during a charity drive at work because he worried that not doing so would make him look stingy.
- b) Wes contributed money to a local charity during a charity drive at work because he believed that doing so would improve his employer's opinion of him.
- c) Donna's employers engaged in an illegal but profitable practice that caused serious damage to the environment. Donna did not report this practice to the authorities, out of fear that her employers would retaliate against her.
- d) Jadine's employers engaged in an illegal but profitable practice that caused serious damage to the environment. Jadine reported this practice to the authorities out of a belief that protecting the environment is always more important than monetary profit.
- e) Leigh's employers engaged in an illegal but profitable practice that caused serious damage to the environment. Leigh reported this practice to the authorities only because several colleagues had been pressuring her to do so.

12. Executive: We recently ran a set of advertisements in the print version of a travel magazine and on that magazine's website. We were unable to get any direct information about consumer response to the print ads. However, we found that consumer response to the ads on the website was much more limited than is typical for website ads. We concluded that consumer response to the print ads was probably below par as well.

The executive's reasoning does which one of the following?

- a) bases a prediction of the intensity of a phenomenon on information about the intensity of that phenomenon's cause
- b) uses information about the typical frequency of events of a general kind to draw a conclusion about the probability of a particular event of that kind
- c) infers a statistical generalization from claims about a large number of specific instances
- d) uses a case in which direct evidence is available to draw a conclusion about an analogous case in which direct evidence is unavailable
- e) bases a prediction about future events on facts about recent comparable events