Which Two Heads are Better than One? Uncovering the Positive Effects of Diversity in Creative Teams

Dutcher, E. Glenn and Rodet, Cortney S.

Ohio University, Ohio University

11 November 2018

Which Two Heads are Better than One? Uncovering the Positive Effects of Diversity in Creative Teams*

E. Glenn Dutcher† Cortney S. Rodet†

November 12, 2018

Abstract

Creative teams drive the idea-economy, yet the determinants of a team's ability to create new ideas are not universally agreed upon. Group-level diversity has gained the most traction as an explanation, where a team's performance is usually attributed to diversity over observed characteristics such as race, gender, or functional expertise. Most agree that these characteristics are not especially important, but rather serve as an indicator of diversity in experiences, which is the actual mechanism that improves team ability. We formalize and test if experientially diverse groups produce more ideas. Because group assignment to projects in the field is rarely exogenous, and experiential diversity is not measured in observational data, we use a laboratory experiment to test our proposal. We find that experientially diverse teams create more ideas and also find no additional effect for gender, racial, socioeconomic, or personality diversity. Our general finding for why diversity may be important indicates that if a correlation exists between characteristic diversity and experiential diversity, the characteristically diverse team will have a higher ability. This generalization can be used to unify divergent results from prior studies and can help explain how dissimilar corporate diversity policies could be equally successful.

JEL Codes: C92, J24, M50, O30, O31, O34

Key Words: Diversity, Creativity, Group Production, Experimental Economics

* The authors thank the members of the Ohio University Research Council for providing funding for this project. Thanks to seminar participants at GATE Lyon St-Etienne, Kent State University, and Ohio University for helpful feedback. Special thanks to Joshua Austin, David Cooper, Daniel Karney, Thomas Ruchti, Tim Salmon, Kirsta Saral, Andrew Smyth, Marie Claire Villeval, and Bart Wilson for very helpful suggestions. Finally, this research would not be possible without the help of Michelle Raines in coordinating laboratory use. † Department of Economics, Ohio University, Athens, Ohio 45701, *email: dutcherg@ohio.edu; rodet@ohio.edu*

1 Introduction

Given the role innovation plays in economic growth, scholars have long sought to understand how new and better ideas are created. One strand of literature posits that existing ideas are combined to form the new ideas that are the driving force behind modern economic growth (e.g., Weitzman 1998). If new creative ideas are the product of one's knowledge and experiences, then greater diversity in knowledge and experiences leads to greater creativity and greater innovation. In problem-solving contexts, it has been shown that teams can be more productive than individuals (e.g., Cooper and Kagel; 2006), and are increasingly used in creativity-centered applications (Jaravel, Petkova, and Bell, 2018; Singh and Fleming, 2010; Wuchty, Jones and Uzzi, 2007; Uzzi et al., 2013). Companies wishing to take advantage of this insight have touted their diversity policies; however, the issue is not settled even among innovative companies.

Consider the dissimilar approaches taken by Bell Labs, originally AT&T's research lab, and Google. Bell Labs established a culture where, "eccentric theorists mingled with hands-on engineers, gnarly mechanics, and businesslike problem-solvers, encouraging the cross-fertilization of theory with engineering" (Isaacson 48). Google's Project Aristotle was meant to understand team effectiveness, where their findings were summed up by one of their analysts as: "We were pretty confident that we'd find the perfect mix of individual traits and skills necessary for a stellar team – take one Rhodes Scholar, two extroverts, one engineer who rocks at Angular JS, and a PhD. Voila. Dream team assembled, right? We were dead wrong." (Rozovsky, 2015)[1] Similarly, Apple's diversity statement claims that "Diverse teams make innovation possible," which is borne out by its diversity in gender and race within the firm. This is counter to a new HR practice of "blind hiring" which asks applicants to strip all identifiers (name, age etc.) from their resumes when applying for a job. Echoing the conflict within organizations, academic studies have found that diversity can lead to group-level conflict (e.g., Putnam 2007) and lower productivity in some instances (Hamilton, Nickerson, and Owan, 2012; Hjort, 2014) and increases in productivity in others (Freeman and Huang, 2015; Lazear, 1999). If diversity is important in generating new ideas, then why is it important, what kind of diversity matters, and what is the underlying mechanism that defines its importance? In response to these questions, we show that diversity matters, and that the type of diversity measured is important. Moreover, we speak to the value of individual experiences at the group level.

[1] https://rework.withgoogle.com/blog/five-keys-to-a-successful-google-team/

When exploring how diversity affects a given outcome, most of the prior literature focuses on race, gender, and socioeconomic background. Prior studies also tend to look at endogenously formed groups. Inherent genetic differences may be present between two or more groups of people, but most scholars would not argue that these genetic differences are why diversity may be important, but rather that the observable characteristics capture other unobserved differences. We posit that the avenue in which diversity leads to an increase in creative capacity, and in turn innovation, is through diverse experiences.[2] This claim implies that groups benefit from members with diverse sets of experiences, which come from social learning, formal learning, and/or belonging to different demographic groups. This feature of our model captures the core intuition for firms' justification that diversity in observable characteristics, such as race, gender, or socioeconomic background, *may* matter because a group which is diverse in these observable characteristics *may* also be diverse in unobservable experiences. To test this core prediction, and clearly establish causal relationships, we report the results of carefully controlled laboratory experiments where we exogenously assigned individuals to groups, measured each group's creative output and analyzed whether this output is affected by the group's diversity of experiences. We also explore whether diversity in observable characteristics, such as race, gender, and socioeconomic background, add explanatory power to our main tests.

The task that most closely matches with our model, and the one that subjects performed in the experiment, is the creative-uses task. In the task, subjects must come up with new and unusual uses for everyday objects, like a cardboard box. Subjects first performed the task on their own before performing it in a group with one other person. This task has been shown to provide a valid measure of divergent thinking, an important component of creativity, and has long been used in creativity research (e.g., Eisengberger, Armeli and Pretz 1998; Torrance 1974). We measured individual subject's breadth and depth of experience or knowledge in thirteen different areas (Carson, Peterson, and Higgins 2005) and combined these individual scores to calculate each groups level of diversity of experiences. We also calculated each group's diversity in demographic characteristics and personality.

First, we find diversity of experiences within a group matters. Increasing the diversity of group experience leads to greater creative productivity, which implies our setting minimizes potential nega-

[2] Anderson et al. (2014) argue that innovation and creativity "being integral parts of essentially the same process." We partially agree, but view innovation as a two stage process where ideas are first created and then implemented where implementation may or may not involve creativity. Our focus is on the idea generation stage.

tive productivity arising from group conflict. Second, we find that, after controlling for experiential diversity, more balanced teams are more productive. That is, diverse groups where one member tends to be far more experienced than the other suffers lower productivity than diverse groups where experience is relatively balanced between group members. Third, we find no additional effects (positive or negative) from group-level diversity in gender, race, socioeconomic background, or personality after controlling for diversity of experiences. Finally, we estimate individual-level and group-level production functions and find that while individual productivity suffers diminishing returns to greater experience, group productivity does not. This final result can be seen as justification for the widespread use of teams in creative organizations.

The benefits of diversity hold great economic potential. Our results point to important factors when forming teams. For instance, building teams with diverse observable characteristics in mind might be fruitful if these characteristics are in fact correlated with the unobservable diversity of experiences. A one standard deviation increase in the level of experience in a group leads to an approximate 12 - 25% increase in creative output. However, when no such correlation between characteristic diverstiy and experiential diversity exists, diversity in observable characteristics might not lead to the desired increase in creative output.

Our results complement existing literature by highlighting why divergent results may have been observed. In some instances, the observable characteristics might have been correlated with the unmeasured diversity in experiences. When this is the case, one can expect a positive effect on productivity. However, when the two are not related, our results imply that diversity in observable characteristics might not lead to the desired result. Our findings that experiences matter, and groups are better able to turn these experiences into output, can be taken as an initial indication that the use of groups in the production process can yield greater economic benefits to a society, however individuals in diverse groups must work well together. Our study can be seen as a showcase environment where group-level conflict is minimized and the promise of diversity in production can be realized. Finally, our results also speak to the broader interest in what determines creativity. Our results show experiences, and not inherent characteristics, are important.

The direct application of the set of questions and tasks we used should not be over-stated. We do not see our study as suggestive that firms start using this questionnaire when forming teams. Rather, the general lessons learned from our study is that first, diversity is important for divergent thinking and second, the reason diversity is important is because of the group members add a meaningful

new perspective.

2 Related Literature

Creativity, problem solving, and team-based production have places in various fields of scholarship including cognitive psychology, management, organizational behavior, and economics. Our investigation bridges these several fields by connecting theory to empirical observation. More specifically, it provides insight into seemingly unsettled findings regarding team-based creative production and diversity among group members.

Creativity research has a long history in cognitive psychology, and although there are many dimensions to creativity, an agreed upon necessary component is divergent thinking (Torrance 1974). Divergent thinking exists when no particular answer is correct, which implies the individual must come up with their own problems to solve (and provide answers to), rather than identifying the single correct answer to a given problem, as is the case with convergent thinking (e.g., an IQ test). Given the ease of administration, and the close tie to definitions of creativity, the creative uses task is frequently used in creativity research as a means of measuring divergent thinking.

Our procedure for measuring participants' experience also links our project to creativity research in cognitive psychology. The Creative Achievement Questionnaire (or CAQ) assesses an individual's knowledge and experience over 13 domains ranging from artistic activities to scientific and intellectual endeavors (Carson, Peterson, and Higgins, 2005). This self-reported measure of experience has been shown to correlate with performance in creative assessments as well as divergent thinking at the individual level (Carson et al., 2005; Kaufman et al., 2016).[3] The CAQ serves as our measure of experience or knowledge that contributes to creative output through analogical transfer.[4] The observed connection between creativity and personality types has important implications for our model with regard to individual choices to gain new experiences. Specifically, self-reported creative achievement, like the CAQ, has been shown to correlate with individuals' openness to new experience (Kaufman et al. 2016). What is unexplored is how different personalities of may work together in the creative domain.[5]

[3]Carson et al. (2005) originally asked subjects to identify those domains in which they have above-average ability or accomplishments and showed responses were externally relevant by providing a link between subjects' responses and a creative activity. We modified the CAQ by asking subjects to rate themselves on a scale of 1 to 7 (four being average) across the 13 domains (see the experiment materials in the appendix).

[4]The results section will explain our measure of diversity using the CAQ.

[5]Diversity in personalities may negatively affect groups, as explained below, so it is one measure of diversity that we

Our project also has obvious connections to brainstorming scholarship that has long sought to understand group dynamics in idea generation (Osborne 1953). Despite its history, the advantages of groups in creative production are still not clear (Kavadias and Sommer 2009; Paulus and Nijstad 2003). Experiments tend to consistently find that the number of ideas generated is greater when individuals work on their own (Diehl and Stroebe 1987; Mullen et al. 1991; Girotra et al. 2010). This is typically explained by free riding, evaluation apprehension (a reticence to express ideas for fear of negative criticism), and production blocking (a natural result of a group's ability to engage one speaker at a time). However, groups tend to outperform individuals when the quality of ideas are taken into account (Girotra et al. 2010). The explanation draws upon a generally accepted theory of creativity in three stages (Singh and Fleming 2010). First is the variation phase, which links directly to our model, where ideas are generated by exploring a combinatorial space of knowledge (Weitzman 1998). The selection phase comes second where ideas are evaluated and promising ones are identified. Thirdly, the retention phase of ideas operates through critical evaluation by the community. Groups bring together a greater combinatorial space of knowledge from which to draw in the variation phase, and intra-group interaction allows for higher quality critique of ideas in the selection phase.[6]

Our research also builds upon scholarship on group diversity. Much like the lack of clarity in comparing groups versus individuals, scholars are still exploring the advantages and disadvantages of group diversity along different dimensions (O'Reilly et al. 1998). Intellectual diversity rooted in education, experience and expertise spurs productivity gains much like the recombinant idea-generation where a more diverse group can tap greater intellectual resources to discover the best approach to problem-solving. Conversely, social category diversity (e.g., gender, ethnic, religious groups) and value diversity, in which members disagree on the objective of the group, inhibit gains from intellectual diversity as interpersonal conflicts over procedures, delegation, and member characteristics arise (Jehn et al. 1999).[7]

evaluate in our analysis. Acemoglu, Akcigit, and Celik (2017) use empirical findings on firm managers and openness to disruption to suggest a causal relationship between openness and innovation using firm-level data.

[6]Singh and Fleming (2010) analyze scientific output and find that research teams are more likely to produce inventions that rank at the 95th percentile or above in terms of citation count (i.e., what they consider to be breakthrough discoveries). They attribute this result to the greater recombinant opportunity in creative search measured by teams' diversity of past inventions (e.g., functional diversity). Furthermore, they attribute a more rigorous selection phase when explaining the significantly lower likelihood of teams to produce inventions that fall into the 5th percentile or below in citation count.

[7]Other scholarship has delved into diversity of cognitive styles to understand their interaction on team creativity (Aggarwal and Woolley 2018; Shalley et al. 2004). Cognitive style refers to how an individual acquires, organizes, and processes information. This has obvious complementarities to the recombinant-idea-generation theory described above; in fact, evidence suggests diverse cognitive styles have positive effects on creative output (Aggarwal and Woolley

Two particularly pertinent papers connecting diversity to problem-solving are by Kavadias and Summer (2009) and LiCalzi and Surucu (2012). Kavaidas and Summer (2009) look at team diversity and brainstorming effectiveness using simulation experiments where agents are programmed according the theoretical model, rather than using laboratory experiments. They find a diverse team tends to generate better solutions, unless the problem is highly complex and requires a high degree of specialization. Importantly, their simulations indicate that teams generate better solutions relative to individuals for problems requiring knowledge from various fields, but diverse teams exhibit an even greater effectiveness. Similarly, LiCalzi and Surucu (2012) model the process of team problem-solving and find in their simulation experiments that two heads can solve a problem even when neither individual could solve the problem on their own if individuals are sufficiently creative.

Our approach to understanding creative teams and diversity utilizes incentivized experiments. Other studies using experiments to understand creativity have primarily focused on incentive structures and their effects on creative output. Motivated by initial findings in social psychology (Amabile, 1996), earlier work on creativity was concerned with questions regarding intrinsic versus extrinsic motivation. More recent work has explored the marginal effects of incentives on creativity, the relative performance of different pay schemes, and the effect that competitive pressure has on creativity. Results are largely mixed where different incentives can increase effort but do not necessarily increase creativity (see Charness and Grieco, 2017; Dutcher, 2012; Ederer and Manso, 2013; Erat and Gneezy, 2015; Gross, 2016; Kachelmeier et al., 2008; Kachelmeier and Williamson, 2010; Kachelmeier et al., 2015).[8] Understanding incentives is extremely important; however, our interest lies in understanding creative problem-solving skills as a function of group experience.

Given the mixed results in the related scholarship, there is an opportunity to better understand creative problem-solving, especially in a team setting. Our focus on diversity of experience, as well as other observable forms of group differences, unpacks the prevailing belief that diversity is a source of creativity and productivity to better understand if it matters and, if so, what type of diversity

2018).

[8]Kachelmeier et al. (2008) and Kachelmeier and Williamson (2010) suggest that pay-for-performance can have immediate impacts on creative output; moreover, pay-for-performance creates gains through self-selection effects. Erat and Gneezy (2015) found that competitive incentives increase effort but not creativity in designing rebus puzzles. On the other hand, Charness and Grieco (2017) found that tournament incentives do not affect creativity in open-ended tasks, but they do increase creative output in tasks with guided direction and constraints. Gross (2016) finds that tournament incentives have an inverted-U-shaped effect on creative performance in designing commercial logos; that is, some competition induces creativity and risk-taking but burdensome tournament incentives can lead participants to drop out of the competition. Ederer and Manso (2013) find that pay-for-performance can induce greater exploration for innovative ideas in a business plan context. Kachelmeier et al. (2015) found that pay-for-performance can have delayed effects on creative output due to incubation effects (Csikszentmihalyi, 2015). When incentives are held fixed, Dutcher (2012) found that an unstructured environment led to higher levels of creative output.

matters.

3 Model

3.1 Overview

Although we spend some time justifying the various elements, the core of the model is quite intuitive: relying on the widely held belief that new knowledge is a function of existing knowledge, it is easy to show that the higher a team's set of existing knowledge, the more new knowledge they can create. When looking at the production of creative ideas, most studies implicitly or explicitly assume production of creative output is achieved by an individual who chooses some costly effort to maximize utility where ability is a given, fixed parameter. Utility is maximized when the marginal benefit is equal to the marginal cost and ability simply scales this this optimal level of effort up (for high ability) or down (for low ability). Prior studies have delved into trying to understand specific portions of this production process which affect the marginal conditions (e.g., Kachelmeir et al. 2008). The domain of these prior incentivized studies can be seen as exploring movement along a production function, while our study can be seen as trying to understand the production technologies that give rise to different production functions. We refer to the production technology as the teams ability where where we argue that diversity within the team determines its ability. Of course, the fundamental question is why one may expect higher ability from a diverse team. Combining insights from various fields, we will outline the premise that the ability of a team is increasing in the number of distinct experiences/knowledge in the team. Although this is the best-case scenario for the team, other literature suggests that diversity will actually lead to a decrease in team production. Distinguishing between the two is left as an empirical question.

3.2 Model of creativity as problem-solving through analogical transfer

The model outlined below synthesizes thinking on creativity as problem solving through analogical transfer (Feinstein 2011; Magee 2005). That is, creativity manifests in discovering novel solutions through the retrieval and application of knowledge and experience, much like Schumpeter's description of the process of economic progress. Cognitive psychologists argue that analogical thinking is fundamental to creativity (See Holyoak and Thagard 1996; Langley and Jones 1988; Schank 1988;

Weisberg 2006). An effective example is Georges de Mestral's discovery of Velcro in 1948 when he noticed how burdock burrs clung to his dog's fur by tiny hooks. His observation of a novel phenomenon – the structure and function of burdock burrs – was paired with what he knew about the relevant use and current limitations of fastening devices to create something new. In other words, through analogical transfer he envisioned how his new discovery could apply to a different setting. Without analogical transfer, the two topics - fastening devices and burdock burrs - would remain separate in his mind.

3.2.1 Components of the creative cognitive process

We organize the multi-level model of knowledge in three levels. At the fundamental level are *elements*, which form principal pieces of knowledge, the combination of which forms new and more complex knowledge. The combinations of distinct elements are labeled as *topics.* For the sake of simplicity, we restrict modeling topics as pairs of elements, but a richer model could allow for greater combinations of elements. More formally, we defined a topic, $t_{1,2} = e_1e_2 = e_2e_1 = t_{2,1}$ $(e_1 \neq e_2)$, as a unique pair of elements. New complex knowledge (i.e. topics) are formed when joining existing elements in a novel pair, combining an existing element with a new (unpaired) element, or pairing two new (unpaired) elements.

By extension, *analogies* are unique pairs of topics, such as $a_{1,2,3} = t_{1,2}t_{1,3}$. We assume two rules about combining topics and their elements into analogies:

Rule 1: Combinations of topics are restricted to pairs.

Rule 2: Two topics can be linked if and only if they share an element.

Therefore, analogies can be specified as $a_{1,2,3} = t_{1,2}t_{1,3} = e_1e_2e_3$. This organization provides some bounded flexibility in that it allows a specific analogy to be formed in three different ways. For instance, $a_{1,2,3}$ can be formed by $t_{1,2}t_{1,3}$, $t_{1,2}t_{2,3}$, or $t_{1,3}t_{2,3}$ where the common elements are 1, 2 and 3, respectively.[9]

[9]If we consider an individual's set of known topics, we can deduce that the overall process of discovering relevant analogies and evaluating their usefulness would involve inordinate amounts of time and insight without some form of constraint on the cognitive process. Put simply, the number of possible analogies increases rapidly with an individual's number of known topics. For this reason, it is important to recognize the rules defined above for pairing topics as constraints.

3.2.2 Analogical transfer function

Our next task is to describe the process by which analogies are constructed to solve problems. This process can also be used to define an individual's potential ability to creatively solve problems by forming novel analogies using the analogical transfer function.

Let a problem be an incomplete pair of topics: $t' = [t_{a,b}, \cdot]$ where $t_{a,b} = e_a e_b$. Myriad solutions to these problems exist where each solution consititutes new knowledge (or creative output). Individual i has the potential to discover t' by a random process $d(E_i)$, where E_i is an individual's personal set of known topics gained through experience, learning, cultural exposure, etc. That is, the likelihood of discovering $t\prime$ is a function of the individual's set of experiences and knowledge. According to our rules, individual $i's$ ability to uncover $t' = [t_{a,b}, \cdot]$ is only possible if E_i contains either: 1) $t_{a,c} = e_a e_c$ or 2) $t_{b,c} = e_b e_c$, where e_c is any element $\neq e_a, e_b$.

The individual's disposition for analogical transfer (i.e. discovering new knowledge by matching a topic with a suitable complementary topic) is the ability to identify useful topic-pairs, or analogies. The specific piece of new knowledge discovered or analogy created is results of the individual's motivation and optimization over some value function, which we briefly address below.[10]

An individual's analogical transfer function can be specified in the spirit of Magee (2005), according to a few fundamental assumptions. First, define the subset of topics in E_i that can be matched with $t\prime$ according to e_a as $E_{i,a}$; likewise define $E_{i,b}$ as the subset of topics that can be matched with $t\prime$ by e_b. We indicate the set of unique topics between these two subsets as $E_{i,a,b} = \{E_{i,a} \cup E_{i,b}\} - \{E_{i,a} \cap E_{i,b}\}$. The number of units within this subset measures an individual's scale of experience, learning and knowledge. Second, each topic in $E_{i,a,b}$ is a potential creative solution to $t\prime$. Third, each topic in $E_{i,a,b}$ has equal potential for being used in analogical transfer (i.e. for simplicity we ignore optimization over $E_{i,a,b}$). Fourth, a topic cannot be paired with itself to solve a problem (this is accounted for in the definition of $E_{i,a,b}$). Fifth, topics within $E_{i,a,b}$ can be paired within an incomplete problem any number of times, which implies that the individual can face $t\prime$ multiple times. We do not put bounds on this number.

Together, these definitions and assumptions produce a function representing an individual's creative problem-solving ability, $s_i(t\prime)$, based on the number of potential solutions they can offer.[11] In other words, $s_i(t\prime)$ describes the number of analogies the individual can form from $E_{i,a,b}$, where

[10] See Sternberg (2006) and Sternberg and Lubart (2006) for an investment theory of creative problem solving.

[11] This does not suggest that all possible solutions that an individual can offer are good or valuable. Identifying the best solution requires a separate, but related, process not included here.

$s_i(t\prime) = f(E_{i,a,b})$.[12] Note that this analogical transfer function is specific to the problem; however, if we were to consider the set of all unsolved problems, call it $T\prime$, we quickly see how this problem-specific function scales up accordingly. The individual's general analogical transfer function approaches a multiple of their entire set of known topics; $s_i(T\prime) = f(E_i)$.

Importantly, problem-solving ability is an increasing function in the number of topics within an individual's E_i. Creative problem solving as a function of experience and knowledge is consistent with existing theories of creativity that rely on a model involving internal filtering, selection and application of ideas (Campbell 1960; Simonton 1999 and 2012), or what Singh and Fleming (2010) describe as the variation phase, the selection phase, and the retention phase. A further assumption that we will make is that E_i is not binary in experience for a given area, but also measures the depth of one's experience in an area where the more experienced an individual is in a given area, the more topics they have at their disposal to solve problems. For instance, consider the number of "economic topics" available to someone who reads an article on economics vs. someone who receives a bachelor's degree in economics vs. someone who receives a PhD in economics.

3.3 Experience

We will briefly address determinants of E_i via the acquisition of different experiences. Doing so will help to give some insights into when diversity in a team can be expected to lead to an increase in creative production and when it may not.

3.3.1 Group Experiences

We will allow an optimal set of experiences, E_i, to be chosen by a utility-maximizing individual given their choice set and preferences, subject to a contraint on their time.[13] E_i captures the magnitude of each experience (how much time was spent on a given experience) and the number of experiences such that $s_i(T')= f(E_i)$ where $s_i(T')$ determines the production technology in a production function.

The same logical process - an increase in experience leads to an increase in ability - also applies

[12] McGee (2005) defines the functional form to be $s_i(T\prime) = w * E_i$ where w represents the number of times a topic can be paired to form an analogy, or an "efficiency" coefficient." For our purposes, we see no reason to define a specific functional form, however, we endorse Magee's (2005) propositions that efficiency reflects cognitive capabilities (e.g. Stiglitz's (1987) concept of learning to learn), depth of knowledge, and motivation for creativity.

[13] In the experiment, we will not observe the individual's preferences or choice sets, but will use a proxy - the CAQ survey which measures their knowledge and skills in thirteen categories - for E_i. Carson, Peterson and Higgins (2005) have shown the external and internal validity of this measure. More details on this measure are given in the Experimental Design section.

to groups. To understand group ability, assume two individuals, i and j, who choose equilibrium experiences $E_i = E_j$ which leads to $s_i(T') = s_j(T')$; the ability to produce creative output for individual i and j respectively given their optimal experiences is the same.[14] If individuals i and j are combined into a group, this group's ability, $s_{i,j}(T')$, is no better than either individual's ability, or $s_{i,j}(T') = s_i(T') = s_j(T')$ which implies that $\{s_i(T') \cup s_j(T')\} - \{s_i(T') \cap s_j(T')\} = 0$, because the joining of member j to a group with member i does not lead to an increase in the set of experiences the group possesses. Now consider individual i being joined with a different individual, k such that $E_i \neq E_k$. Because the two individuals have different experiences, this implies $s_{i,k}(T') > s_i(T')$, $s_{i,k}(T') > s_k(T')$, and $\{s_i(T') \cup s_k(T')\} - \{s_i(T') \cap s_k(T')\} > 0$.[15] We will call the latter team diverse and the former team homogeneous. Because diverse teams have more topics available to them, the analogical transfer function dictates that they will be able to solve more problems: i.e., $s_{i,k}(T') > s_{i,j}(T')$.

Hypothesis 1: Creative output is increasing in a team's set of diverse experiences.

The main point is that a team is more productive iff the two members on the team have a different set of experiences. There are several reasons two individuals may differ in their chosen experiences. The most obvious is when individuals have different preferences. If these individuals face the same choice set and the same constraint, these different preferences will lead to different optimal choices. For instance, prior studies have highlighted some instances where the preferences of males differ from those of females (Croson and Gneezy, 2009). If the identified preferences are important in optimal experience selection, the chosen experiences for males and females will differ and pairing a male with a female will lead to an experientially diverse group, which we would hypothesize is more productive than a non-diverse group.[16] Another reason individuals may make different choices could be due to differing choice sets. For instance, most females are not allowed to (or are socially discouraged from) participating in male-centric sports and visa-versa. Similarly, those in a higher social class may be

[14] For instance, if i and j have the same preferences, the same choice sets and the same constraint, they will choose the same experiences.

[15] For simplicity, consider an instance where $s_i(T') = s_k(T')$. One way for this to occur is if the preference ordering of experiences by individual k were the inverse of i, which would lead to $E_i \neq E_k$. Similarly, E_i and E_k could be singletons if these individuals have very strong preferences for one experience. However if these preferences differ, $E_i \neq E_k$ and, assuming both face the same constraint, $s_i(T') = s_k(T')$. One can easily construct other preference relationships where a similar result can be obtained.

[16] The qualifier is important here and highlights that gender may or may not be a good indicator when putting together a diverse team.

discouraged from certain "low-class" experiences while those in a lower social class may not be able to afford some "high-class" experiences. Thus, even if two individuals have the same preferences, we will observe different optimal choices if their choice sets differ. If the group's experiences are appropriately controlled for, we should observe no additional effect for observationally (or otherwise) diverse teams.

Hypothesis 2: After controlling for experiential diversity, characteristic diversity will not have an addtional effect on creative output.

Even though our structure suggests a positive effect from diversity, numerous studies previously mentioned highlight that such an assumption may not hold.

Hypothesis 3: An experientially or characteristically diverse team will be less productive.

3.4 Analogical transfer in the Creative-uses task

We argue that the creative-uses task, explained in more detail in the next section, is an appropriate test of creativity through analogical transfer, because it requires subjects to take what they know about a particular object and identify ways it can be used to solve problems for which it was not intended. Given an object that has functional characteristics and properties represented by $t_{a,b}$, subjects are tasked with identifying different t' for which $t_{a,b}$ serves as an appropriate match, i.e. to form appropriate and useful analogies. To do so, subjects mentally search for many problems to which an object's functional characteristics and properties can be applied; nevertheless, we contend that the cognitive processes are the same and that greater experience will increase an individual's or group's ability to form those analogies.

4 Experimental design and procedures

This section explains our experimental design, including details regarding the creative-uses task, as well as our collection of subject characteristics used to measure group diversity.

4.1 Creative uses task

In the experiment, subjects performed the creative-uses task (CUT). To perform the task, subjects were given the name of an object with a short description of its intended use. Subjects then had seven minutes to come up with as many alternative uses for the object as they could. For instance, if the subjects were given the object "tin can", a creative use would be to use the tin can as a flower pot, cut it up and use the pieces as the blades of a pinwheel or put a string between two tin cans and use as a toy telephone. They performed this task four times where each time, they were given a different object. The objects were a brick, cardboard box, metal pipe, and t-shirt.

Submissions were judged for their validity and uniqueness by two independent research assistants. The judges deemed a use valid if they believed the use could reasonably be implemented while uniqueness was according to the submissions within the same session. Feedback was not immediate due to the effort needed to judge submissions, so subjects knew that their payment would be delayed about a week while our independent judges calculated their payments.

4.2 Questionnaire

Subjects were also asked to complete a questionnaire that provided measurements of group diversity. The first part of the questionnaire asked demographic questions about subjects' age, race, native English speaker, gender, and socioeconomic background. The last item was measured on a three item scale where subjects were asked if their family's income while growing up would be considered below average, average, or above average.

Next, subjects completed the Creative Achievement Questionnaire (CAQ) which features 13 distinct areas of achievement including: visual arts, music, dance, theater and film, architectural design, creative writing, humor, culinary arts, individual sports, team sports, entrepreneurship, scientific inquiry, and inventions. The typical use of the CAQ asks subjects to indicate the areas in which they believe they are above average (Carson et al. 2005). We modified the CAQ by asking subjects to rate themselves in each area on a 7-point scale indicative of their talent, ability, or training, where 4 meant average achievement, 1 indicated that the subject was way below average on that item while a score of 7 indicated that they were way above average. We implemented this richer measure to better capture magnitudes, which allowed us to better estimate the number of "topics" available to each individual. This formulation also allows us to better measure relative

experiences/knowledge in those instances where many subjects may believe they are above or below average on a given topic. If the binary approach were used, we would miss out on these relative comparisons which, according to our model, may be important.

Lastly, subjects completed a short 10-item personality assessment with a 7-point scale indicative of their agreement with seeing themselves as: extroverted and enthusiastic; critical or quarrelsome; dependable and self-disciplined; anxious and easily upset; open to new experiences; reserved and quiet; sympathetic and warm; disorganized and careless; clam and emotionally stable; conventional and uncreative. Creativity research has established correlations between individuals' openness to new experiences and self-reported creative achievement (Kaufmann 2016). Furthermore, diversity in personality potentially diminishes group creative production by introducing conflict according to 'process' conflict (O'Reilly 1998).

4.3 Experimental procedures

When subjects arrived to the experimental session, they were randomly assigned a seat. Once subjects were seated at their workstation, instructions were handed out to all participants and read out loud to ensure common knowledge. After all questions were answered, the first period began. Subjects performed the CUT four times. Subjects saw the word on their screen and were provided a space where they could type their creative use. As soon as they clicked the button to accept their entry, the entry appeared below their typing box and they were then allowed to type another use. An on-screen timer was used so subjects would know how much time was remaining.

In the first period, they worked alone for seven minutes and their pay was based entirely on their own output. Participation in the first period familiarized the subjects with the task. Subjects were paid \$0.15 for every valid submission and another \$0.15 for each unique (and valid) submission, as judged by our research assistants. In periods 2-4, subjects sat side-by-side with their partner and submitted creative uses into a single computer terminal. Because they were allowed to communicate verbally, we employed white-noise to reduce inter-group influence. Each subject's pay was based on the aggregation of both member's output. Each group member was paid \$0.15 for every valid submission and \$0.15 for every unique submission that the pair came up with. Subjects were free to leave after completing the questionnaire and were reminded that we would contact them within a week with instructions on where to pick up their payment. Sample instructions are found in the appendix.

5 Results

Six experimental sessions were conducted. We recruited subjects online and conducted the experiments using z-Tree (Fischbacher 2007). In total, we had sixty-six subjects, and thirty-three two-person groups.

5.1 Summary statistics

We use two measures of performance in the task: 1) the total number of submissions per group per object as the product of the variation-phase effect and 2) the absolute valid output per group per object as the product of the selection phase. Because we had two independent judges, our measure of valid output is the average number of valid uses per subject/group per object.[17] There are at least two reasons to use the number of uses submitted as a measure of performance. First, even though our judges agreed substantially regarding the validity of submitted uses, the agreement between them was not perfect. This a natural result of discretionary evaluation. The number of submitted uses does not suffer from this issue. Second, some techniques for creative problem solving implicitly rely on the number of ideas generated as a measure of productivity (see Higgins (1994) for examples).

Variable		Variable	
Number of Subjects	66	Below Average Income	0.17
			(0.38)
Number of Groups	33	Average Income	0.50
			(0.50)
Submissions (period 1)	16.82	White	0.79
	(5.61)		(0.41)
Valid (period 1)	13.91	African American	0.06
	(4.34)		(0.24)
Submissions (p. 2 - 4)	26.39	Asian	0.05
	(6.75)		(0.21)
Valid (p. 2 - 4)	22.94	Other	0.11
	(5.74)		(0.29)
Male	0.55	Age	22.02
	(0.50)		(4.14)

Table 1: Overview of the data. When appropriate, means and standard deviations (in parentheses) are provided.

Table 1 contains the summary statistics according to demographic variables. Table 2 displays

[17] We had decent agreement from our two judges on the validity of the subjects inputs where the degree of correlation was 0.87 ($p<0.01$) and Cronbachs alpha was estimated to be 0.92.

Key Variables	Mean (St. Dev.)
Ind. experience	52.42 (10.01)
Group experience	64.42 (7.42)
Balance	0.60 (0.28)
All Male	0.33 (0.48)
All Female	0.24 (0.43)
Female/Male	0.42 (0.50)
Race	0.36 (0.49)
Income	0.64 (0.49)
Inv. HHI Personality	1.24 (0.10)

Table 2: Group-level summary statistics for diversity variables

group-level summary statistics of our measures of diversity across different subject characteristics. These measures are explained in the next sections.

5.2 Diversity in experiences

We now move on to our primary research question: how does diversity affect creative output? We will first focus on diversity in experiences and then move to diversity in other measures. We use the CAQ to measure the diversity of experiences.

We take several approaches in exploring the question of diversity of experiences. First, the group's experience is calculated using the group-specific maximum score within each CAQ category based on the assumption that the most experienced in the group within a given category will influence creative problem solving along that dimension because a higher CAQ score implies more experience. That is:

$$Gexp_g = max\left[caq_{1g}^1, caq_{2g}^1\right] + max\left[caq_{1g}^2, caq_{2g}^2\right] + \cdots + max\left[caq_{1g}^{13}, caq_{2g}^{13}\right] \quad (1)$$

This variable increases with group-level experience by treating every level in each CAQ category as a distinct topic (in the model's nomenclature) and could have a value between 13 and 91. It is clear that combining two individuals who have the same score in a particular CAQ category does not lead to any gains in experience. Our goal is to test if raising the $Gexp$ by combining individuals

who are different leads to productivity gains.[18] This variable allows us to measure diversity within a given team along two dimensions. First, increases in *Gexp* can be attributed to an increase in the maximum score in any category, meaning it captures intra-category diversity between team members. Second, *Gexp* can increase by *adding* an additional category to the experience set that would not be there if the individual were to work alone. This captures inter-category diversity between team members.

The variable *Gexp* could be a result of a single individual on a team who is very experienced, or it could be due to a diverse team. To clearly establish the effects of diversity, we also include estimates from a model that includes two variables in addition to *Gexp* that decompose indivdiual ability. The first is the maximum individual summation of CAQ scores across all categories in each group, or what we call $iMax$. This controls for the individual contribution of the most experienced individual in a very simple way (although it leaves the reader with a slightly more challenging interpretation of the results given the ceteris paribus nature of regression analysis). The second variable is the minimum individual summation of CAQ scores, or what we call $iMin$.

Likewise, we include estimates from a third specification that includes a variable that controls for the contributions made to *Gexp* by each team member. Therefore, we constructed a variable called *Balance* that captures the balance of a groups contributions to overall experience. That is:

$$Balance_g = 1 - \frac{countmax_{ig} - countequal_g}{13 - countequal_g} \tag{2}$$

where $countmax_{ig}$ is the count of CAQ categories of the greatest individual contributor of the group (where i references the individual and g references the group), and $countequal_g$ is the count of CAQ categories in which the group members have identical scores. This variable was further normalized to vary between $[0, 1]$, where 0 indicates that *Gexp* is completely based on a single individual's CAQ scores, and 1 indicates that no group member contributes more to *Gexp* than the other. This variable performs a function similar to including $iMax$ and $iMin$ to some extent, but we argue that it more carefully controls for individual contribution and has a more intuitive interpretation.[19]

[18] To check if this specification was correct, we also explored if non-diversity enhancing experiences affected output. In Table 7 of the Appendix, we include a variable Gmin which accounts for the level of experiences not accounted for in Gexp. That is, this variable is the sum of the group-level minimum of the experiences in each category. From the table, it is shown that this variable is not statistically significant and does not alter the main results on Gexp and balance.

[19] It is important to address the distinction between our approach and the common interchanging of diversity and dispersion as a statistical measure of difference or variance. Firstly, diversity of experience as dispersion is not specifically what our model implies. Rather, our focus is on the added specialization that diverse experience brings to a group. It is not that because two people are different that makes them more creative as a team, it is that together

	Submitted			Valid output		
Regressors	1	2	3	1	2	3
Gexp	0.42***	0.89***	0.49***	0.31***	0.83***	0.35***
	(0.15)	(0.24)	(0.13)	(0.12)	(0.22)	(0.10)
iMax		-0.52 **			-0.52**	
		(0.26)			(0.22)	
iMin		0.14			0.03	
		(0.14)			0.12	
Balance			6.69**			4.37*
			(3.22)			(2.51)
% English	7.94	6.63	6.82**	7.86	6.72*	7.13**
	(5.24)	(5.02)	(5.01)	(3.75)	(3.44)	(3.58)
Constant	-8.32	-12.77	-15.55	-3.70	-6.75	-8.42
	(13.63)	(11.61)	(11.71)	(10.37)	(8.82)	(8.78)
Wald Chi2	48.18	70.61	58.19	26.36	69.47	37.17
R-sq:						
Within	0.38	0.38	0.38	0.27	0.27	0.27
Between	0.30	0.42	0.39	0.28	0.40	0.34
Overall	0.32	0.41	0.38	0.28	0.36	0.32
Obs.			99			
Groups			33			

Table 3: The marginal effect of diverse experiences on submitted output and valid output from panel random-effects regression; standard errors clustered at group level; period dummies also included; ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

We regressed submitted output and valid submitted output at the group level against *Gexp*, *iMax*, *iMin*, *Balance* and other controls, including the proportion of group members who are native English speakers, and indicators for the specific objects in the CUT. Standard errors are clustered at the group level.[20]

Table 3 contains the regression results.[21] Columns 1 and 4 indicate the results when we do not control for balance. We see that increasing group experience increases submitted and valid uses. For context, a one-unit increase in *Gexp* (reminder, *Gexp* could range from 13 to 91) increases submissions by 0.42 units and valid submissions by 0.31 units.

Consider the standard deviation of the group-level experience shown in Table 2, which is 7.42 points, for context. A one-standard-deviation increase in group-level experience translates into 3.12 more submissions overall and 2.30 more valid submissions in a 7-minute period. The average per-

each person adds experience to the group that the other does not. Together, our *Gexp* and *Balance* variables work to capture these differences. We elaborate on this point in the discussion section.

[20]We considered allowing for non-linear effects from *Gexp*, *iMax*, *Balance*, and their interaction and found that under such a functional form, the marginal effect of each of these variables is essentially linear. Average marginal effect estimates are essentially the same as those in Table 3.

[21]To check for robustness of these results, in Table 6 of the Appendix, we include a session by session analysis and find results are largely consistent with the pooled sample.

period number of submissions was 26.39 and the average per-period number of valid submissions was 22.94. This would be equivalent to increases of 12% and 10%, respectively.

When decomposing the individual contributions to group-level experience, the second specification indicates that the marginal effects due to *Gexp* increases to 0.89 for submissions and 0.83 for valid submissions. This suggests increases of 6.60 submissions and 6.16 more valid submissions, for a one-standard-deviation increase in group experience (i.e., increases of 25% and 27%, respectively)

When controlling for balance, the third specification indicates that the marginal effects due to *Gexp* increases to 0.49 for submissions and 0.35 for valid submissions. This suggests increases of 3.64 more submissions and 2.60 more valid submissions, for a one-standard-deviation increase in group experience (i.e., increases of 14% and 11%, respectively.) This is in line with Hypothesis 1 and against Hypothesis 3.

Result 1: Greater diversity of experience and knowledge within a team leads to increases in both the number of creative uses submitted and the number of valid uses submitted.

The marginal effect of $iMax$ is also significant in both the submission and valid output specifications, but it is negative. Given the ceteris paribus nature of the analysis, the variable indicates that if two groups have the same *Gexp* score, the group with a lower iMax score will be more productive. A similar interpretation can be seen from the balance variable.

The marginal effect of balance is also significant. As *Balance* increases from 0 to 1, meaning a change from complete imbalance to complete balance, submissions increase by 6.69 and valid submissions increase by 4.37. This suggests increases of 25% and 19%, respectively.

Result 2: Greater balance in individual contributions to group-level experiential diversity, whether measured by $iMax$ or by Balance, increases submissions and valid submissions.

5.3 Diversity in Gender, Race, Income, and Personality

We now explore if diversity in gender, race, income and personality provide additional explanatory power. We first outline how we define diversity in these measures and then show analysis of these measures on output.

Gender: We use a simple measure of gender diversity where we use indicators for groups that are all-male, all-female, or female/male. In our sample, 42% of groups were mixed gender.

Race: We condensed our racial categories to five: White, African American, American Indian, Asian, and Other. The Other category incorporates small subgroups in our sample. We constructed an indicator variable indicating whether group members are of the same racial group or not, without concern for the specific mixes of racial groups. That is, the indicator is 1 if group members are of different racial groups and 0 otherwise. In our sample, 38% were mixed-race groups.

Income: Our measure of background socioeconomic status simplified the measure of subjects socioeconomic background by indicating whether family income while growing up was below average, average, or above average. This variable is again, just a binary variable indicating that the group members are or are not from the same socioeconomic background. Overall, 64% of groups were socioeconomically diverse.

Personality: We assessed personality along 10 dimensions, each with a scale of 1-7. We constructed an inverse Herfendahl-Herschman Index for personality using individual indicators for when individuals rated themselves above 4 in a given category. This gives us an effective number of types according to the multiple personality dimensions. A value of one implies that group members are above or below average in all of the same personality categories; a value of two implies that group members are on different sides of the mean on every personality category. Values between 1 and 2 imply the group members are similar in some categories and different in others.

We regressed output against *Gexp*, *Balance*, and the gender, race, income, and personality diversity measures, separately. The same control variables were included, and standard errors are clustered at the group level.

Table 4 reports the marginal effects of the key diversity measures in each regression. The left side of the table includes the estimates where the dependent variable is the number of submitted uses. The right side contains the estimates where the dependent variable is the number of valid submissions. The effect of *Gexp* and *Balance* remains robust to the inclusion of these other controls

	Submitted				Valid output			
Regressors	1	2	3	4	1	2	3	4
Gexp	0.46***	0.48***	0.49***	0.49***	0.33***	0.35***	0.36***	0.36***
	(0.12)	(0.13)	(0.14)	(0.13)	(0.10)	(0.10)	(0.11)	(0.10)
Balance	7.14**	6.39**	6.53*	6.70**	5.01*	4.12*	4.16*	4.34*
	(3.46)	(3.01)	(3.43)	(3.23)	(2.71)	(2.38)	(2.58)	(2.54)
All Male	-0.11				0.23			
	(1.77)				(1.31)			
Female/Male	-1.80				-1.84			
	(1.87)				(1.60)			
Race		-0.88				-0.76		
		(1.42)				(1.77)		
Income			-0.56				-0.75	
			(2.06)				(1.77)	
Inv. HHI Personality				-0.42				1.06
				(7.20)				(6.04)
% English	6.04	6.82	6.93	6.83	6.31	7.12**	7.27*	7.12*
	(5.07)	(5.06)	(5.39)	(5.06)	(3.73)	(3.58)	(3.99)	(3.59)
Constant	-12.75	-13.91	-15.00	-14.90	-5.63	-7.02	-7.68	-10.09
	(11.43)	(12.63)	(11.23)	(14.03)	(9.07)	(9.59)	(8.29)	(10.63)
Wald Chi2	58.08	60.32	65.44	59.72	45.03	37.71	49.25	43.07
R-sq:								
Within	0.38	0.38	0.38	0.38	0.27	0.27	0.27	0.27
Between	0.41	0.39	0.39	0.39	0.38	0.34	0.34	0.34
Overall	0.40	0.39	0.39	0.38	0.35	0.32	0.32	0.32
Obs.				99				
Groups				33				

Table 4: The marginal effect of demographic diversity on submitted output and valid output from panel random-effects regression; standard errors clustered at group level; period dummies also included; ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

in significance and magnitude. That is, greater diversity in experiences increases submissions and valid submissions, as does a balanced contribution to group experience by both group members.[22]

We find that neither gender, racial diversity, socioeconomic diversity, nor personality diversity has additional robust explanatory power. Taken as a whole, we do not find strong support for others measures of diversity adding much explanatory power.

Result 3: Diversity of gender, race, socioeconomic background, and personality do not have a measurable effect on creative production when accounting for diversity of experience.

[22]From this point on, we use the *Balance* variable to control for individual contribution to *Gexp* because of its interpretation. It is highly correlated with $iMax$ because they essentially measure the same thing. Results using $iMax$ are nearly identical throughout and are available upon request.

5.4 Comparing individual and group production functions

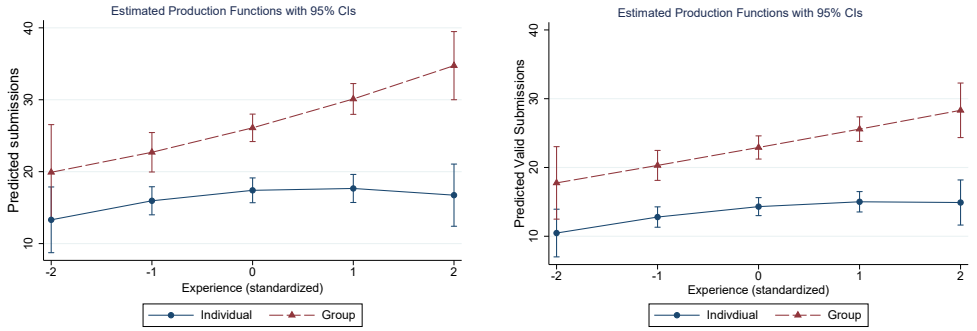
The manner in which individuals and teams are able to utilize their experiences may differ due to constraints at the individual level, such as imperfect recall (of experiences) or biases inherent with an increased cognitive load. This may imply that as the number of experiences increases, an individual may have an increasingly harder time accessing these experiences, while a team may not. Exploring this possibility allows us to address why firms frequently use team collaboration in creative problem solving, rather than relying on independent output of numerous individuals. One potential benefit of working in groups is not only the diversity of experiences but also the mitigation of diminishing returns to an expanding experience set.

For example, let one individual have an experience set equal to $E_i = \{t_a, t_b, t_c, t_d, t_e, t_f, \}$ and two individuals in a group have the respective experience sets equal to $E_j = \{t_a, t_b, t_c\}$ and $E_k = \{t_d, t_e, t_f, \}$. Because the individual and the group both have the same experience set, we may conclude that they produce the same output - $s_i(t') = s_{j,k}(t')$ - if diminishing returns do not set in. Now consider a production function that exhibits declining marginal returns at the individual level, $h(s(t'))$ such that $h' > 0$ and $h'' < 0$. If the group production is the sum of the individual's ability, a team will be more productive than an individual with the same experience given the members of the team would be able to jointly "access" more experiences than the individual can on their own. More generally, if declining marginal returns are present in this setting, in order for an individual to be as productive as a team, the individual must have an experience set such that $E_i = \alpha(E_j + E_k)$ where $\alpha > 1$ depends upon the degree of declining marginal returns.[23] The implication is that group problem solving potentially pushes the production frontier out (relative to individuals with the same experience) because groups may not suffer diminishing returns to experience. We explore the validity of this claim empirically.

We estimated simple production functions for submissions and valid submissions at the individual and group level. The individual production function is estimated using the first period output when subjects worked alone and incentives were based on individual output. The group production function is estimated using a modified random-effects panel regression that controls for *Gexp* and *Balance* and allows for non-linear interaction within and between the *Gexp* and *Balance* variables.

[23] An explicit example would be $h_i(s(t')) = \sqrt{s_i(t')}$ and $h_g(s_g(t')) = \sqrt{s_j(t')} + \sqrt{s_k(t')}$. Given the number of elements in the respective experience sets in the example, the individual with the larger experience set would produce approximately 2.44 analogies, while the group could produce approximately 3.46, even though each individual in the group has fewer experiences than the first individual.

Figure 1: Individual and Group-level Production Functions



The individual level regression uses the individuals CAQ average and allows for non-linear effects by including a quadratic term.

Figure 1 displays the estimated relationship between experience and creative output at the individual and group levels. The left panel shows the predicted submissions over the standardized sample ranges of individual CAQ and $Gexp$.

The data suggest that individual output exhibits diminishing returns to increasing experience whether we consider submissions or valid submissions. On the other hand, output at the group level is strictly increasing in experience. For individuals, submissions are estimated to be approximately 13 for an individual with experience equal to 2σ below the mean. Meanwhile, an individual with average experience is estimated to submit approximately 17 uses. The number of submissions stays roughly the same for an individual that is 2σ above the mean in experience. Groups with experience 2σ below the mean produce 20 submissions, whereas groups with average experience produce nearly 26 uses. Groups with experience 2σ above the mean produce nearly 35 submissions. Results are similar for valid submissions where similar estimates range from 11 valid uses to 15 valid uses for individuals, and estimates range from approximately 18 to 28 valid uses for groups.

Result 4: Estimated production functions indicate that individuals encounter decreasing returns to experience; however, groups exhibit constant returns to experience.

6 Discussion

We address our results regarding diversity in observable characteristics and our concept of diversity in experiences in this section.

6.1 Generalization of diversity, experiences, and production

In this section, we outline a general framework for analyzing when diversity in observable characteristics may lead to higher team-level productivity. In our outline, we will assume that the incentives to work together are strong, as is the case in our study, though there certainly exists instances where diverse group members may not work well together.[24] The foundation of this exploration still relies on the experience sets of the individuals in the team as the main driver of productivity.

From the set of n experiences available to an individual, let a subset of these experiences, $m \subset n$, be key to complete some task, k. That is, the most productive two-person team at some task k is the one which has jointly chosen their optimal experiences such that $E_i + E_j = m$ where E_i and E_j are the optimal sets of experiences defined previously. When determining if observable characteristics, such as gender, race and/or socioeconomic background, affect productivity for task k, one must define the set of m experiences that are needed and assess if a correlation exists between m and the observable characteristics. In cases where such a correlation exists, we can expect a team composed of demographically diverse individuals to perform better. However, when a correlation is absent, then forming a team based on observable characteristics may not lead to the desired increase in productivity. As was previously stated, the differences in the optimal experiences arise from either a difference in preferences or constraints on an individual's choice set where the constraints can be socially or otherwise determined.

Take for instance the results of the current study. In our study, we utilized the CUT which is an accepted proxy for measuring divergent thinking. Thus, our results apply when the task, k, relies heavily on divergent thinking. Additionally, we have shown that a relevant m for this setting can be gathered via a CAQ questionnaire. Our population was college students who were roughly the same age. It is not hard to imagine that, in the current environment, they may have all had similar choice sets available to them when they chose their optimal set of experiences. At least, that has

[24] See, for instance, Chen and Li (2009) for an overview of how group identity is often used in social situations such as teamwork where incentices, such as those identified in Bandiera et al. (2009), can be the catalyst for negative effects arising from group identity.

been the hope for diversity policies over the last decades. If, on the other hand, we had conducted the same experiment in a previous decade or in a country with fewer diversity policies (and/or more discrimination), increases in production of demographically diverse groups may have been found. However, the result would still rely on the experiences of the individuals.

6.2 Measurement of group diversity

We now address the definition of diversity and how it can be measured. As a reminder, we are measuring the ability of a group according to:

$$Gexp_g = max\left[caq_{1g}^1, caq_{2g}^1\right] + max\left[caq_{1g}^2, caq_{2g}^2\right] + \cdots + max\left[caq_{1g}^{13}, caq_{2g}^{13}\right] \tag{3}$$

The implication of this measurement is that the only way for individual members to add to the group's ability is if the two individuals differ on which categories they are better at. If this occurs, then the group has more "topics" which leads to a greater ability to solve problems. This definition fits our descriptive model. For instance, let there be two groups where group 1 is composed of individual 1 and 2 and group 2 is composed of individual 3 and 4. For simplicity, assume only two categories are used. Let $E_1 = [2, 4]$, $E_2 = [4, 2]$, $E_3 = [6, 7]$ and $E_4 = [7, 6]$, such that $Gexp_1 = 8$ and $Gexp_2 = 14$ where E_i denotes individual $i's$ experience set and $Gexp_d$ denotes group $d's$ ability. Because $Gexp_1 < Gexp_2$, group 2 is predicted to be more productive than group 1. Furthermore, because each member in both groups is better than the other in different categories, both group benefits from each member's experience. That is, because the members are *different* from each other and these differences add to the group's collective experience set, the group is better from their inclusion.

In other fields, such as biology, ecology, or finance (portfolio theory), diversity often implies dispersion, or how different each individual is, should be the measurement used for diversity. To see why measuring diversity as dispersion may not make sense in our setting, let us introduce a third group composed of individuals 5 and 6 where $E_5 = [1, 1]$ and $E_6 = [3, 4]$ which implies that $Gexp_3 = 7$. With our measurement of ability, we would predict that group 2 is more productive than group 1 which is more productive than group 3. Let us introduce a definition of dispersion which is equal to:

$$Gdisp_g = |caq^1_{1g} - caq^1_{2g}| + |caq^2_{1g} - caq^2_{2g}| + \cdots + |caq^{13}_{1g} - caq^{13}_{2g}| \tag{4}$$

That is, this measure captures how different each individual is on the team without regard for what that team member may be bringing to the team in terms of their *added* experiences. From this new measure, $Gdisp_3 = 5 > Gdisp_1 = 4 > Gdisp_2 = 2$. If $Gdisp_g$ is used as a measure of ability, we would arrive at completely different predictions of which team will be the most productive. We show in the appendix that using this measure of dispersion leads to inferior estimation. More generally, it is difficult to come up with a logical framework that fits with the current literature on creativity that would justify using $Gdisp$ as a measure of diversity. In general, we believe that using such a measure leads to incorrect inference as to the effect of diversity and $Gexp_g$ should be used in the domain of creative production. It should also be pointed out that there will likely exist a correlation between $Gdisp_g$ and $Gexp_g$ because $Gexp_g$ can capture differences between individuals, however $Gdisp_g$ also captures components of differences between the individuals on a team which do not have any logically consistent basis.

7 Conclusion

Ample evidence suggests that diversity may be societally harmful. More ethnically diverse societies in Sub-Saharian Africa pass inferior public policies (Easterly and Levine, 1997) and, in the U.S., Putnam (2007) found that in ethnically diverse neighborhoods, trust, cooperation and altruism were lower. These negative effects are typically tied to inter-group conflicts, a result predicted by social identity theory (e.g., Akerlof and Kranton, 2000 and Tajfel and Turner, 1979) and partially supported by Chen and Li (2009). Though diversity may entail some costs, diversity in a society and in a firm is thought to be beneficial via increased productivity (Lazear, 1999, Alesina and La Ferrara, 2005), however, there exists very little support for this notion. Hamilton, Nickerson and Owan (2012) showed that productivity was lower for diverse teams in a garment factory. Hjort (2014) finds that ethnic diversity leads to lower productivity in a flower factory. Functional diversity has been shown to hold the greatest promise of a positive effect; however Berge, Juniwaty and Sekei (2017), Lyons (2017) and de Oliveira and Nisbett (2017) call into question the general applicability of such a finding. Taken as a whole, a diverse group faces a seemingly insurmountable hurdle in increasing production. However, these domains may not have given diversity its best chance of succeeding. If

diversity is to play a positive role, it is often hypothesized that this role will be through increased creativity - a key ingredient in innovation. Even here though, the evidence is mixed and many policies focus on characteristic diversity such as gender, race, or personality. We confirm that the underlying rationale for why these characteristics may matter - group-level diversity in knowledge and experiences leads to more creative output. We find no additional effect from characteristic diversity.

The final result highlights that if characteristic diversity is found to be important, it is likely important because of different experiences and not because of genetic differences. When might we expect characteristic diversity to increase creative output? We argue it may be when preferences differ between individuals from characteristically diverse groups or when the choice sets available to these two groups differ (with some assumptions on how the sets do or do not overlap). In such an instance, these differences lead members of the groups to choose a different set of optimal experiences: i.e., experientially diverse groups can be formed from characteristically diverse groups.

It should be mentioned that our results need not apply to every task which may be described as a "creative" task. We have simply identified the set of experiences where diversity within a team matters for our task (divergent thinking). Should a different task be used, a potentially different set of experiences may need to be identified. If that is the case, it may very well be that, due to different choice sets/preferences, characteristically diverse teams are a feasible way to form an experientially diverse team. However, once again, the key is experiential diversity.

We also find that after controlling for the group's total experience, more balanced teams, in terms of the experience each individual possesses, are more productive than less balanced teams. Thus, our model does not completely capture some aspects of a team's ability, but because a group's total experience is found to positively affect output, even after controls for individual contributions are included, implies that it is diversity within the team that is important. Moreover, putting diverse groups to work avoids diminishing returns to knowledge and experience observed at the individual level. This implies that whatever promise diversity holds for an increase in innovation, it is going to be realized to a greater extent when diverse individuals interact in a team. Mechanisms which reduce group-level conflict are key to unlocking diversity's potential.

We provide a clear path for diversity to increase performance, and our study should be viewed as such. Having established this path, future studies can systematically build upon this to understand how other factors, such as limited communication, group size and group turnover, mitigate or enhance

the effects found in our study. Doing so allows a deeper understanding of the processes needed to take advantage of increasingly diverse societies.

8 References

Acemoglu, D., Akcigit, U., & Celik, M. A. (2014). Young, restless and creative: Openness to disruption and creative innovations (No. w19894). National Bureau of Economic Research.

Aggarwal, I., & Woolley, A. W. (2018). Team Creativity, Cognition, and Cognitive Style Diversity. Management Science.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. The Quarterly Journal of Economics, 115(3), 715-753.

Alesina, A., & Ferrara, E. L. (2005). Ethnic diversity and economic performance. Journal of economic literature, 43(3), 762-800.

Amabile, T. M. (1996). Creativity in context: Update to the social psychology of creativity. Hachette UK.

Anderson, N., Potonik, K., & Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. Journal of management, 40(5), 1297-1333.

Bandiera, O., Barankay, I., & Rasul, I. (2009). Social connections and incentives in the workplace: Evidence from personnel data. Econometrica, 77(4), 1047-1094.

Berge, L. I. O., Juniwaty, K. S., & Sekei, L. H. (2016). Gender composition and group dynamics: Evidence from a laboratory experiment with microfinance clients. Journal of Economic Behavior & Organization, 131, 1-20.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. Psychological review, 67(6), 380.

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. Creativity Research Journal, 17(1), 37-50.

Charness, G., & Grieco, D. (2017). Creativity and financial incentives. University of California, Santa Barbara, Working Paper.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. American Economic Review, 99(1), 431-57.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. Journal of Economic literature, 47(2), 448-74.

Csikszentmihalyi, M. (2015). The systems model of creativity: The collected works of Mihaly Csikszentmihalyi. Springer.

de Oliveira, S., & Nisbett, R. E. (2017). Culture changes how we think about thinking: From Human Inference to Geography of Thought. Perspectives on Psychological Science, 12(5), 782-790.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. Journal of personality and social psychology, 53(3), 497.

Dutcher, E. G. (2012). The effects of telecommuting on productivity: An experimental examination. The role of dull and creative tasks. Journal of Economic Behavior & Organization, 84(1), 355-363.

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. The quarterly journal of economics, 112(4), 1203-1250.

Ederer, F., & Manso, G. (2013). Is pay for performance detrimental to innovation?. Management Science, 59(7), 1496-1513.

Eisenberger, R., Armeli, S., & Pretz, J. (1998). Can the promise of reward increase creativity?. Journal of Personality and Social Psychology, 74(3), 704.

Erat, S., & Gneezy, U. (2016). Incentives for creativity. Experimental Economics, 19(2), 269-280.

Feinstein, J. S. (2011). Optimal learning patterns for creativity generation in a field. American Economic Review, 101(3), 227-32.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental economics, 10(2), 171-178.

Fleming, L. (2001). Recombinant uncertainty in technological search. Management science, 47(1), 117-132.

Freeman, R. B., & Huang, W. (2015). Collaborating with people like me: Ethnic coauthorship within the United States. Journal of Labor Economics, 33(S1), S289-S318.

Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea generation and the quality of the best idea. Management science, 56(4), 591-605.

Gross, Daniel P. "Creativity Under Fire: The Effects of Competition on Creative Production." Harvard Business School Working Paper, No. 16-109, March 2016. (Revised September 2016.)

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2012). Diversity and productivity in production teams. In *Advances in the Economic Analysis of participatory and Labor-managed Firms* (pp. 99-138). Emerald Group Publishing Limited.

Hjort, J. (2014). Ethnic divisions and production in firms. The Quarterly Journal of Economics, 129(4), 1899-1946.

Holyoak, K. J., Holyoak, K. J., & Thagard, P. (1996). Mental leaps: Analogy in creative thought. MIT press.

Isaacson, W. (2014). The Innovators: How a Group of Inventors, Hackers, Geniuses and Geeks Created the Digital Revolution. Simon and Schuster.

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.

Kachelmeier, S. J., Reichert, B. E., & Williamson, M. G. (2008). Measuring and motivating quantity, creativity, or both. Journal of Accounting Research, 46(2), 341-373.

Kachelmeier, S. J., Wang, L. W., & Williamson, M. G. (2015). The Role of Incentives in Sustaining High-Creativity Production. Working Paper, University of Illinois at Urbana-Champaign and the University of Texas at Austin. Available at HBS: www. hbs.edu/faculty/Lists/Events/Attachments/942/Williamson_creativity_3_4_15. pdf.

Kachelmeier, S. J., & Williamson, M. G. (2010). Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity. The Accounting Review, 85(5), 1669-1691.

Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B., & DeYoung, C. G. (2016). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. Journal of Personality, 84(2), 248-258.

Kavadias, S., & Sommer, S. C. (2009). The effects of problem structure and team diversity on brainstorming effectiveness. Management Science, 55(12), 1899-1913.

Langley, P., & Jones, R. (1988). A computational model of scientific insight. The nature of creativity: Contemporary psychological perspectives, 177, 201.

Lazear, E. P. (1999). Culture and language. Journal of political Economy, 107(S6), S95-S126.

LiCalzi, M., & Surucu, O. (2012). The power of diversity over large solution spaces. Management Science, 58(7), 1408-1421.

Lyons, E. (2017). Team production in international labor markets: Experimental evidence from the field. American Economic Journal: Applied Economics, 9(3), 70-104.

Magee, G. B. (2005). Rethinking invention: cognition and the economics of technological creativity. Journal of Economic Behavior & Organization, 57(1), 29-48.

MacKay, D. J., & Mac Kay, D. J. (2003). Information theory, inference and learning algorithms. Cambridge university press.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. Journal of personality and social psychology, 52(6), 1258.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. Basic and applied social psychology, 12(1), 3-23.

O'Reilly III, C. A., Williams, K. Y., & Barsade, S. (1998). Group demography and innovation: Does diversity help?.

Osborne, A. F. (1953). Applied imagination: principles and procedures of creative problem solving. Charles Scribener's Sons, New York.

Paulus, P. B., & Nijstad, B. A. (Eds.). (2003). Group creativity: Innovation through collaboration. Oxford University Press.

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twentyfirst century the 2006 Johan Skytte Prize Lecture. Scandinavian political studies, 30(2), 137-174.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. Creativity Research Journal, 24(1), 92-96.

Schank, R. C. (1988). Creativity as a mechanical process. The nature of creativity, 220-238.

Schumpeter, J., & Backhaus, U. (2003). The theory of economic development. In Joseph Alois Schumpeter (pp. 61-116). Springer, Boston, MA.

Shalley, C. E., Zhou, J., & Oldham, G. R. (2004). The effects of personal and contextual characteristics on creativity: Where should we go from here?. Journal of management, 30(6), 933-958.

Simonton, D. K. (1999). Origins of genius: Darwinian perspectives on creativity. Oxford University Press.

Simonton, D. K. (2012). Taking the US Patent Office criteria seriously: A quantitative three-criterion creativity definition and its implications. Creativity research journal, 24(2-3), 97-106.

Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality?. Management science, 56(1), 41-56.

Stiglitz, J. E. (1987). Learning to learn, localized learning and technological progress. Economic policy and technological performance, 125-153.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. The social psychology of intergroup relations, 33(47), 74.

Torrance, E. P. (1974). The Torrance Tests of Creative Thinking. Norms-Technical Manual Research Edition. Verbal Tests, Forms A and B- Figural Tests, Forms A and B. Princeton, NJ: Personnel Press.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. Science, 342(6157), 468-472.

Weisberg, R. W. (2006). Creativity: Understanding innovation in problem solving, science, invention, and the arts. John Wiley & Sons.

Weitzman, M. L. (1998). Recombinant growth. The Quarterly Journal of Economics, 113(2), 331-360.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. Science, 316(5827), 1036-1039.

9 Appendix

9.1 Experimental Instructions

Welcome to today's experiment. I would like to thank you for participating. In the course of today's experiment, you will have a chance to earn money based on your decisions. All decisions and answers will remain confidential and anonymous. All monetary amounts you will see in this experiment are denominated in standard dollars and cents. You will be paid your cumulative earnings over all of your decisions along with a participation fee of $5 that you were promised to receive regardless of the outcome of today's experiment.

Before we begin, it is extremely important that you put away all materials including external reading material, pens and pencils and turn off your cell phones and any other electronic devices. If you have a question, please raise your hand and I will come by and answer it privately. Also, please do not talk to each other or try to look at the screens of others during the experiment.

Today's experiment consists of several parts. We will give you instructions for each part separately.

First, notice that in your instructions packet, we have given you a sheet of paper with your name and a personal ID code. We have a duplicate of this paper. Since we do not link your name to the data, this ID code is used to identify you. Once you collect your payment, this sheet will be shredded so that there is not a link between your choices and your name. You can enter your id on your computer now.

Part 1

In the first part you will be given seven minutes to come up with alternative uses for an object displayed on your screen. The uses you come up with must be different than the use for which the object was originally designed or intended. As an example, suppose the object that appeared was "a tin can." Some uses which you would get credit for could be a "flower pot", a "holder for my dreams", "attach a string between two cans to make telephones", "cut the can and make a pinwheel", etc. Intended uses which you would not get credit for, would involve things like "storing carrots" or "container for evaporated milk" since these are uses the tin can was intended for. Also, you would not get credit for uses that are the same. In the example above, if you had put "a daisy pot" and "a petunia pot" this would get counted as one valid use since both are considered flower pots. You will enter each use in the space provided on the screen and hit the "Next" button to record it. Once you hit the "Next" button, your input will be recorded and show up beside the input box at which point you can type another use. The inputs you enter must not be longer than 100 characters and you are limited to a total of 40 inputs. You will see the remaining time on your computer (in the top right corner). When the time is up, you will not be able to enter new words. If you turn to your computer screen, we will show you how the computer interface works.

Payment:

For this task, you will receive $0.15 for each use you come up with in which the object was not intended for and an additional $0.15 for each use that no one else in the group participating in this session comes up with. Thus, for each use there is the potential to make $0.30. Let's go through an example. Assume you typed 20 uses for the object which satisfy the above criteria and earned $0.15 for each or $3.00. Also, assume you came up with 5 uses which are unique to this session and earned an additional $0.15 for each or $0.75. In this hypothetical example, your total payment would be $3.75.

The validity and uniqueness of your responses can not be judged immediately and so you will not see immediate feedback on your earnings. Your responses will be recorded and judged by two independent judges (not the experimenters) to determine whether or not they satisfy the requirements. This process will be completed within one school week (excluding holidays or planned days off at the University). You will be notified by email when the process is completed. In the email, you will be given instructions on how to collect your payment.

Are there any questions? If not, click on the "Continue" button and we will proceed.

Parts 2, 3 & 4:

Parts 2, 3 and 4 of the experiment are similar to Part 1; however, you have now been paired with another subject with whom you will work for the next part of the experiment. You and your partner will be given a different object in each Part and asked to come up with alternative uses as a partnership. Only one of you will type uses into the computer. You may communicate with each other during this part of the experiment, but please keep your voices down as not to influence other participants. We will be using white-noise to help maintain privacy between partnerships. Because only one of you will type uses into the computer, there will be only one active screen for each group. The first screen you see will inform you if your screen is the active one, and if not, where to move to be next to your group member. Once again, the uses you come up with must be different than the use for which the object was originally designed or intended. You will again have seven minutes in each Part to come up with as many uses as possible. The payment for each use judged as valid is $0.15 and the payment for each use that is unique is $0.15 for a potential payment of $0.30 per valid entry. The validity and uniqueness of the uses will be judged by two independent judges.

Are there any questions? If not, click on the "Continue" button and we will proceed. If your screen indicates to move to your partner's computer, please do so now.

Part 5

In Part 5, you will be presented with the title of, and a short paragraph about, a recent movie. Your task is to list alternative titles for the movie. Once again, only one of your screens will be active. You may communicate with each other during this part of the experiment, but please keep your voices down as not to influence other participants. You will have seven minutes to come up with alternative titles. Two

independent judges will determine if a given title seems feasible and if it is unique. Unique is as previously defined, while feasible titles are those which the judges deem as a reasonable title a movie producer would use if given the description. For each feasible title you come up with, you and your partner will receive $0.15. For each unique title, you and your partner will receive an additional $0.15. You will not get credit for titles which are too similar. As an example, suppose the movie was "White Fang," a film about a young gold prospector and his wolfdog. The alternative titles must be feasible titles given the movie description; however, note that similar feasible titles will only receive credit once. For example, if you had put "a boy and his dog looking for gold," "a boy and his wolfdog," and "a gold prospector and his dog," these would get counted as one valid title since all are close variations of each other.

Are there any questions?

Part 6:

In Part 6, you will have a final opportunity to earn more money, by choosing to play one of five lotteries. This decision will be made individually rather than in pairs. The outcome of each lottery is determined from a virtual coin flip. Please choose which one of the coin flip lottery games you would like to play. You can only pick one. After you have marked your choice, click "Flip" to determine the outcome (your payoff). Heads and Tails are equally likely.

Part 7:

1. What is your gender?
2. What is your age?
3. What is your major?
4. Is English your native language?
5. What country are you originally from?
6. To the best of your ability, indicate the income level that best describes your family's annual household income:
   ___ Below Average
   ___ Average
   ___ Above Average

Personality Assessment (Brief Big Five; Gosling, Rentfrow and Swann 2003)

Here are a number of personality traits that may or may not apply to you. Please indicate a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Disagree strongly =1; Disagree moderately = 2; Disagree a little = 3; Neither agree nor disagree = 4; Agree a little = 5; Agree moderately = 6; Agree strongly =7

I see myself as:

1. ____ Extraverted, enthusiastic
2. ____ Critical, quarrelsome
3. ____ Dependable, self-disciplined
4. ____ Anxious, easily upset
5. ____ Open to new experiences, complex
6. ____ Reserved, quiet
7. ____ Sympathetic, warm
8. ____ Disorganized, careless
9. ____ Calm, emotionally stable
10. ____ Conventional, uncreative

Brief Creative Achievement Questionnaire (Carson, Peterson and Higgins 2005)

Please indicate in the left column what areas you have been repeatedly exposed to either through a class or as a hobby. Please indicate in the right column what areas you feel you have more talent, ability, or training than the average person by checking the box next to it.

__ visual arts __

__ music __

__ dance __

__ individual sports (tennis, golf) __

__ team sports __

__ architectural design __

__ entrepreneurial ventures __

__ creative writing __

__ humor __

__ inventions __

__ scientific inquiry __

__ theater and film __

__ culinary arts __

9.2 Instructions to judges

Your task is to decide if the uses that the subjects came up with fit the criteria of an unintended use. Instructions subjects saw (in italics):

In the first part you will be given five minutes to come up with alternative uses for an object displayed on your screen. The uses you come up with must be different than the use for which the object was originally designed or intended. As an example, suppose the object that appeared was a tin can. Some uses which you would get credit for could be a flower pot, a holder for my dreams, attach a string between two cans to make telephones, cut the can and make a pinwheel, etc. Intended uses which you would not get credit for,

would involve things like storing carrots or container for evaporated milk since these are uses the tin can was intended for. Also, you would not get credit for uses that are the same. In the example above, if you had put "a daisy pot" and "a petunia pot" this would get counted as one valid use since both are flower pots. You will enter each use in the space provided on the screen and hit the Next button to record it. Once you hit the Next button, your input will be recorded and show up beside the input box at which point you can type another use. The inputs you enter must not be longer than 100 characters and you are limited to a total of 40 inputs. You will see the remaining time on your computer (in the top right corner). When the time is up, you will not be able to enter new words. If you turn to your computer screen, we will show you how the computer interface works.

In the excel spreadsheet, which has the output of each subject (which is tied to their randomly determined subject number), place an x in the box if the use they typed in fits the criteria of valid. If there is no possible way the object can be used in the manner given, it does not get credit. The stated use does not have to be sensible; it only needs to be feasible. Below are definitions for the words they were given and attempted to come up with unusual uses for.

> **brick** : a handy-sized unit of building or paving material typically being rectangular and about 2.25 3.75 8 inches (57 95 203 millimeters) and of moist clay hardened by heat used as a building or paving material.
>
> **cardboard box**: a rigid typically rectangular container with or without a cover Used to store or transport goods.
>
> **extension cord**: an electric cord fitted with a plug at one end and a receptacle at the other used to power electronic devices.
>
> **metal pipe**: a long tube or hollow body made of metal - used for conducting a liquid, gas, or finely divided solid or for structural purposes.
>
> **t-shirt**: a collarless short-sleeved or sleeveless shirt usually made from cotton used as clothing.

Below is a list of example uses for a flower pot which **would** be counted as valid.

1. Paint it and use it as art
2. Hold water
3. use as a pan to cook with
4. Hat
5. Make
6. sand castles
7. Serve deserts in it
8. Turn it upside down and use the whole at the bottom as a flag prop
9. Smash and use broken bits as a knife

10. keep in in your car floor as a trash can
11. prop up a window
12. use it to trap bugs
13. use it as a bird bath
14. use it as a stepping stool
15. Cut holes for eyes and make a mask out of it
16. Use it as a shower caddy
17. Use it for target practice
18. use it as a stool
19. a holder for my dreams
20. use it as a toilet
21. pour candle wax inside and make it into a candle

Below is a list of examples for a flower pot which **would** not be counted as valid.

1. Plant a daffodil in it
2. Plant a tomato plant
3. Plant a cucumber plant in it
4. Plant a sunflower in it
5. Plant a cactus in it
6. Plant a plastic flower in it
7. Plant a pepper plant in it
8. Plant basil in it
9. Plant oregano in it
10. Plant cilantro in it
11. Plant a lily in it
12. Plant a garlic bulb in it
13. Plant a shrub in it
14. Plant aloe vera in it
15. Plant agave in it
16. Plant zinnias in it
17. Plant a rose bush in it
18. Plant a fern in it
19. Plant ficus in it
20. Plant tulips in it

9.3 Checking robustness of results

This short section contains the baseline regression results on diversity of experiences estimated by session to show that the effects from *Gexp* and *Balance* are fairly robust, even at the low numbers in each session.

	Submissions						Valid output					
Regressors	1	2	3	4	5	6	1	2	3	4	5	6
Gexp	0.76***	0.64***	0.85***	0.56***	0.62***	0.37	0.83***	0.53***	0.62***	0.46***	0.52***	0.10
	(0.35)	(0.14)	(0.28)	(0.04)	(0.15)	(0.42)	(0.39)	(0.13)	(0.35)	(0.05)	(0.08)	(0.24)
Balance	-7.58	10.86***	-17.10	5.04***	13.13***	4.86	-4.76	6.70***	-17.14	4.64***	10.37***	4.30
	(14.04)	(1.59)	(8.83)	(1.11)	(2.42)	(7.48)	(15.54)	(1.40)	(10.87)	(1.40)	(1.11)	(4.14)
R-sq:												
Within	0.63	0.47	0.43	0.49	0.67	0.16	0.29	0.45	0.29	0.57	0.80	0.04
Between	0.78	0.91	0.88	0.98	0.84	0.28	0.70	0.87	0.77	0.96	0.92	0.37
Overall	0.74	0.73	0.82	0.82	0.78	0.25	0.59	0.67	0.70	0.80	0.86	0.24
Obs.	18	18	15	15	15	18	18	18	15	15	15	18
Groups	6	6	5	5	5	6	6	6	5	5	5	6

Table 5: Regression analysis of diversity of experience by experimental session. Random-effects panel regression with clustered standard errors.

The results below include an additional measure of skill that captures non-diversity-enhancing ability. This additional measure is the sum of the *minimum CAQ* scores across group members. It does not provide additional explanation of creative output.

	Submitted		Valid output	
Regressors	1	2	4	5
Gexp	0.39**	0.51***	0.33***	0.42***
	(0.15)	(0.11)	(0.13)	(0.09)
Gmin	0.06	-0.04	-0.04	-0.12
	(0.17)	(0.15)	(0.14)	(0.13)
Balance		6.99**		5.22**
		(3.18)		(2.58)
% English	7.97	6.75	7.84**	6.93**
	(5.28)	(5.03)	(3.79)	(3.52)
Wald Chi2	47.90	72.74	27.82	66.31
R-sq:				
Within	0.38	0.38	0.27	0.27
Between	0.30	0.39	0.29	0.36
Overall	0.32	0.39	0.28	0.33
Obs.		99		
Groups		33		

Table 6: Regression analysis including non-diversity-enhancing skill. Random-effects panel regression with clustered standard errors; ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

The results displayed below show that when testing for fit using the encompassing method (Davidson and MacKinnon, 2003), the baseline model outperforms other models using other measures of experience diversity.

The encompassing method essentially makes a nested model from non-nested models. Wald tests are used to test the restrictions that the additional measures are redundant. In all four cases, we cannot reject the null hypothesis that the additional control for diversity of experience is redundant.

	Submitted					Valid output				
Regressors	1	2	3	4	5	1	2	3	4	5
Gexp	0.49***		0.55***		0.47***	0.35***		0.54***		0.31**
	(0.13)		(0.20)		(0.18)	(0.10)		(0.17)		(0.14)
Gavg		5.86**	-1.07				3.77**	-2.92		
		(2.30)	(3.72)				(1.92)	(3.28)		
Gdisp				3.17**	0.53				3.20***	1.46
				(1.39)	(1.86)				(1.19)	(1.64)
Balance	6.69**	3.90	7.03**	6.02	7.03**	4.37*	2.29	5.32	4.66	5.32**
	(3.22)	(3.23)	(3.17)	(3.92)	(3.17)	(2.51)	(2.64)	(2.53)	(3.06)	(2.53)
Wald Test p-value			0.77		0.77			0.37		0.37
Wald Chi2	58.19	52.18	70.89	52.09	70.89	37.17	31.49	68.05	46.40	68.05
R-sq:										
Within	0.38	0.38	0.38	0.38	0.38	0.27	0.27	0.27	0.27	0.27
Between	0.39	0.29	0.39	0.13	0.39	0.34	0.22	0.36	0.19	0.36
Overall	0.38	0.31	0.39	0.19	0.39	0.32	0.24	0.33	0.22	0.33
Obs.	99									
Groups	33									

Table 7: Encompassing method test of fit using competing measures of diversity of experience; $Gavg$ = group average CAQ score; $Gdisp$ = sum of absolute difference in CAQ category scores; Wald Test null hypotheses refer to testing competing measures of diversity of experience; ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.