Problem Solving by Heterogeneous Agents

Lu Hong*and Scott E. Page†‡

February 17, 2000

Abstract

A substantial amount of economic activity involves problem solving, yet economics has few, if any, formal models to address how agents of limited abilities find good solutions to difficult problems. In this paper, we construct a model of heterogeneous agents of bounded abilities and analyze their individual and collective performance. By heterogeneity, we mean differences in how individuals represent problems internally, their *perspectives*, and in the algorithms they use to generate solutions, their *heuristics.* We find that while a collection of bounded but diverse agents can locate optimal solutions to difficult problems, problem solving firms can exhibit arbitrary marginal returns to problem solvers, and that the order that problem solvers are applied to a problem can matter, so that the standard story of decreasing returns to scale may not apply to problem solving firms.

1 Introduction

In this paper, we analzye problem solving activities in the economy from a microeconomic perspective. Broadly speaking, by problem solving, we mean finding solutions that generate revenues, so this includes activities such as searching for cures to diseases, developing software, handling legal cases, and constructing social policies. We contrast problem solving with other productive activities such as manufacturing by explicitly modeling how agents individually and collectively search for improving solutions. We model both how an agent encodes a problem and how she generates potential solutions. Based on this model, we are able to establish that in the absense

*Department of Economics, Syracuse University, Syracuse, NY 13244. lhong@syr.edu

†Department of Economics, University of Iowa, Iowa City, IA, 52242. scotte@page.biz.uiowa.edu

‡The authors would like to thank seminar participants at Caltech, The Santa Fe Institute, Northwestern, UC–Irvine, Rochester, Queen's, Iowa, NBER-NSF Decentralization Conference, and the Econometrica Society Winter Meetings in San Francisco. They would also like to thank Stan Reiter, John Ledyard, Ehud Kalai, Matt Jackson, Jenna Bednar, Jerry Kelly, and Tom Nohel for comments on earlier versions of this paper. They would also like to thank two anonymous referees and an associate editor for their careful reviews which have greatly enhanced the exposition of this paper.

of incentive and communication problems, diverse perspectives and/or approaches of agents often lead to optimal solutions to complex problems, despite the fact that each individual agent may not be able to locate a good solution if acting alone. This result provides an intuitive and plausible explanation for an increasingly popular phenomenon that high tech products are often developed on an open source by a loosely defined group of people.[1] We also show that as long as problem solvers are not so intelligent as to be able to locate the optimal solution single handedly or they are not so ignorant that they have no knowledge about the problem they are solving, marginal returns to problem solvers do not follow a particular pattern of increasing, decreasing, or constant. This result suggests that insights gained through general equilibrium models based on the assumption of decreasing marginal returns often do not apply to problem solving activities.

Problem solving has become a large part of the economy. Knowledge workers, which includes people who solve problems and people who process information, comprise the largest segement of the United States' work force. The range of human capital involved in the production and classification of knowledge is diverse, yet economics has, by and large, treated workers as having unidimensional ability, reducing humans to robots of various speeds or memory. Implicitly, this assumption requires either that people do not differ in our behavior, some of us are just more or less intelligent than others, or that economically relevant activities require workers to perform standard and repetitive tasks. Forty years ago, when a majority of workers produced goods and services (when they made stuff), models assuming a unidimensional ability parameter may be beneficial. They yielded insights and proved accurate empirically. Many economists do not know that the Cobb–Douglas production function was originally an empirical result.[2] In the context of general equilibrium theory, assuming identical workers performing repeated, standard tasks led to tractable models of production, and more broadly speaking, implications that were theoretically and empirically pleasing: wage equals marginal productivity.

In the modern economy, a minority of people earn their incomes by engaging in the manufacture of goods. In the United States and other advanced economies, more people provide services, process information (Radner 1993), and solve problems than manufacture goods. Many commodities that are manufactored such as CDs and software, have huge fixed costs and small marginal costs. It stands to reason then that the canonical general equilibrium model of economic activity might benefit from amending, or at a minimum from a reinterpretation of its foundations. While we do not deny that in many cases, particularly for the provision of services, the neoclassic model of Arrow-Debreu (1954), Debreu (1959) and McKenzie (1959) remains an accurate representation, in other cases, such as information processing (Radner and Van Zandt 1995), the application may be less secure. In this paper, we explore the

[1] For example, the Linux Operation System.

[2] See Douglas (1967) for an interesting account of the early development of Cobb-Douglas production function.

extent to which the standard story applies to firms that solve problems.

To our knowledge, no models exist that analyze firms that solve problems, firms that engineer new drugs, develop software, design homes and bridges. Workers engaging in these activities perform nonstandard tasks and human ingenuity plays an important role in their performance. This observation fundamentally differentiates problem solving from other production activities, specifically manufacturing, where workers perform standard tasks. More importantly, it leads us to explicitly model economic agents with distinct understandings of the world and varying approaches to solving problems – agents whose differences go beyond non-identical priors over states of the world; agents whose richness of problem solving techniques mirrors that which we see in the world. A problem solving model of that type must be tractable enough to allow for comparisons of agents on common problems and flexible enough to handle groups of agents solving problems.

Any model of problem solving must relax the optimality assumption. Although our agents possess intelligence, they cannot always locate the optimal solution to the problem. No one yet knows how to cure the common cold, or fold proteins, let alone construct an efficient, fair educational system. To most people, this assumption seems reasonable. One novelty of our model lies in that we assume problem solvers differ in how they encode and approach problems. This not only captures human individuality but also provides the basis for an explanation of why collections of agents outperform individuals: by virtue of being different, individuals can improve upon each other's solutions to a problem. In this first model, we assume that every solution to the problem can be evaluated and that all problem solvers agree on the value. For simplicity, we can interpret the value as the amount of money the solution would sell for in the market.

In the formal model, each problem solver is characterized by a *perspective/heuristic* pair: a *perspective* is a problem solver's internal representation of a problem, an encoding; a *heuristic* is an algorithm, or rule(s) of thumb that a problem solver applies in searching for a solution. The final solution that a problem solver locates depends on the interplay of her perspective and heuristic. Problem solvers may differ along either dimension or along both.

Though our explicit characterization of agents as perspective/heuristic pairs is new, [3] it is not without foundation. In their book *Human Problem Solving*, Newell and Simon (1972) modeled the human thought process as following two steps: representation of the problem and application of a heuristic, although not necessarily only once in the process. Such artificial intelligence models are not without critics (see Penrose 1989), but many such models have made their way into economics in subfields as diverse as macroeconomics (Sargent 1993), game theory (Kalai and Lehrer 1993 & 1995), and political economy (Kollman, Miller, and Page 1993). In microe-

[3]It has recently been brought to our attention that de la Maza and Yuret (1994) mentioned the idea of explicitly modelling the perspectives and heuristics (they use the terms "model sets" and "algorithm sets" respectively) of traders in futures markets, though no explicit modelling was present in their paper.

conomics, the relevance of agents following rules of thumb to the theory of firm has been argued since at least as early as the work of March and Simon (1958) and Cyert and March (1963). Artificial agents have been defined as automata (Rubinstein 1986 and Kalai and Stanford 1988), perceptrons (Cho 1993), genetic algorithms (Arifovic 1994, Miller 1992), and classifier systems (Marimon, McGrattan, and Sargent 1990). In these models, agents choose heuristics given a fixed perspective which is chosen by the modeler. Our framework generalizes many of these models.

Our model extends artificial agent models by allowing agents to have different perspectives. Adding this dimension permits a richer description of human capital and enlarges the scope of diversity in problem solving. By including perspectives, we can distinguish between the knowledge embedded in the framing of a problem and the knowledge represented by the cognitive tools used to solve the problem. We show through the construction of equivalence classes of problem solving behavior that perspectives *truly* enlarge the set of ways of solving a problem as compared to if the perspective is fixed and only heuristics vary. Two agents belong to the same equivalence class if their perspective/heuristic pairs are indistinguishable in how they locate solutions to a problem. We also establish a lower bound on the number of equivalence classes. The lower bound becomes astronomically large as the problem domain increases. This result indicates the possibility of even greater diversity in problem solving for a less restrictive environment which resonates with our motivation for group problem solving.

Focusing on a binary string model of collective problem solving, we derive two other categories of results. The first demonstrates the possibility of collective optimality despite limited individual ability. We find that diversity in either perspectives or heuristics proves sufficient for a collection of agents to locate optimal solutions to difficult problems. We want to point out here that these findings apply equally well to a collection of agents over a long period of time working on the same problem. **So, our model could be applied to describe the historic sequence of performance improvements in a particular product such as the combustion engine or the airplane.** Nothing in our analysis requires that the agents make a decision as a group. All of the agents may work in isolation. We need only assume that they have a common value function, **an assumption that need not be true in political settings**.

The second category of results relies on interpreting our model as representative of problem solving within firms. We analyze the returns to additional workers, the improvement in the value of the best solution brought about by adding a worker. We arrive at some unexpected findings. We demonstrate the possibility of arbitrary returns to additional problem solvers, and show that an identical group of problem solvers applied to a problem might exhibit increasing returns or decreasing returns depending upon the order they are hired, even if each agent has "equal problem solving ability". Finally, we derive sufficient conditions for the returns to additional problem solvers to be decreasing. The conditions describe situations where problem solvers do not have particular insights to the problem they are solving. Such situations clearly defy economic logic: Economic agents choose problems for which they have skill, as opposed to choosing to attack problems that confound them. Taken together, our results suggest less regularity in the returns to additional workers for problem solving firms than for other types of firms. Our richer description of human capital in problem solving allows a problem solver's marginal contribution to be context dependent: An agent's marginal product depends upon the relationship between her human capital and those of the other problem solvers.

Before proceeding with our analysis, we should mention several features not included in our model. We ignore asymmetric and imperfect information since our purpose is to construct a model of economic agents with distinct understandings of the world that go beyond information issues. We also ignore incentives. This assumption may be problematic given the importance placed on incentive constraints in organizational structure and performance (Milgrom and Roberts 1992). Relaxing this assumption could introduce differences in the values of objects, which opens the door to preference cycles and agenda manipulation. Finally, though we ignore communication problems – solutions can be costlessly and errorlessly communicated to other agents – our use of perspectives should make a formal model of communication costs tractable. We hope to address this in future work.

We have organized the remainder of this paper in six parts and two appendices. In section 2, we present data describing the number of problem solving workers in the U.S. economy and discuss our assumptions about heterogeneity in perspectives and heuristics in greater detail. We find the number of problem solving workers to be substantial — not too different from the number of manufacturing workers at present — and increasing. In section 3, we present a formal mechanistic model of heterogeneous individual problem solvers based upon the perspective/heuristic dichotomy.[4] We focus on the special case in which agents rely on the same mathematical language for their perspectives: binary strings, and construct a formal measure of diversity. In section 4, we prove a link between diversity among bounded agents and collective optimality. In section 5, we analyze the returns to adding problem solvers. We find the possibility of arbitrary returns. We also find that a reordering of agents of "equal ability" can shift the returns to adding problem solvers from increasing to decreasing. In section 6, we establish the aforementioned equivalence relationship among perspective/heuristic pairs. We conclude with a discussion of the robustness of our results as well as comment on extensions including problem solvers with different value functions, miscommunication between problem solvers, and distinctions between humans and computers as problem solvers.

2 Background

This paper rests on two background assumptions: (1) advanced economies contain a sufficient percentage of people whose work consists of solving problems to warrant

[4]See Marr (1982) for a definition of mechanistic theories of behavior.

creating a separate theoretical apparatus and (2) diverse perspectives and heuristics at least partially explain the benefits of collections of agents to outperform individuals at solving problems. We present data from the United States to support the first assumption. The second assumption seems non controversial. Fresh perspectives and new ideas provide a basis for many improvements in problem solving contexts.

2.1 Problem Solvers: The Data

Before entering into a micro–level discussion of job classifications, we report some aggregate statistics which support our first assumption. In the 1950's when general equilibrium models were developed, the U.S. economy was nearly one half manufacturing. At present, less than one–fifth of the U.S. economy is classified as manufacturing. In fact, more people currently work for state, local, and federal governments than work for manufacturing firms. Most workers in developed economies do not produce goods. They either provide services, process information, or solve problems. Blurry lines distinguish these three classifications, so assigning workers to each requires great care. Radner (1992) has estimated that up to forty percent of U.S. workers are information processors, but his estimate includes workers one could also classify as problem solvers.

One approach to estimating the number of problem solvers in an economy would be to count the number of firms solving problems. Accepting this definition, the problem solving segment would include many technological firms, such as software development companies, much of the entertainment industry, including movie production companies, and many professional firms, such as consulting companies and law firms.[5] And in fact, classifying firms in this way, we find that the portion of the economy, whether measured in GDP or number of workers, comprised of problem solving firms has grown over the past forty years.

An alternative approach, and the one undertaken here, relies on data on job classifications. We count the number of workers whose job descriptions suggest that they spend a substantial amount of time solving problems. This approach includes workers employed by firms whose primary purpose is to process information, manufacture goods, or provide services but which contain problem solving subunits. These subunits may formulate strategies, provide legal defense, determine incentives, or contemplate the restructuring of the organization. In addition, this approach excludes workers who though employed by firms which solve problems, actually provide services, process information or manufacture goods. The data presented are taken from *The Statistical Abstract of the United States* and *The Statistical History of the United States*. The table below summarizes those workers whose primary responsibility is problem solving. Determining which workers solve problems and which process information is more art than science. We have chosen to include management consultants, lawyers, medical researchers, and computer programmers as problem solvers

[5]In their book, *The Winner Take All Society,* Frank and Cook present data showing huge increases in size for the top consulting firms.

but to exclude bankers and elementary school teachers. Undoubtedly, this approach leads to some mistakes. Equity analysts, bankers who do solve problems, will not be counted. Moreover, in that many manufacturing workers now spend part of their time participating in problem solving, we are certainly underestimating the extent that modern workers devote to solving problems.[6] Nevertheless, our crude estimates reveal an unmistakable trend.

The Increase in Problem Solvers [7]

Year	**1950**	**1970**	**1992**
problem solvers (1,000)	6,081	9,630	21,392
total work force (1,000)	59,230	79,802	117,598
percentage of total work force	10.3	12.1	18.2

In sum, these data show both a substantial increase in the percentage of problem solvers in the economy over the last forty years as well as the significant portion of the modern economy devoted to problem solving.

2.2 Improvement Through Diversity

Unfortunately, we cannot produce such overwhelming hard data in support of our second assumption, that heterogeneity of perspectives and heuristics explains the benefits of collections of agents solving problems. In the literature, there are several reasons put forth as to why groups outperform individuals. The theory presented here encompasses many of them. Given the theoretical focus of this paper, we refrain from summarizing existing theories of group performance, but we will touch on one example. Consider the specialization explanation for group performance. This says that groups are better at solving multidimensional problems because individuals can develop expertise on components of the larger problem. Our model accounts for this phenomenon – agents' heuristics can apply to only a portion of the domain, thereby mimicking specialization.

The modest goal of this subsection is to promote the simplicity and plausibility of our second assumption: given the crucial caveat that all individuals agree on the value of outcomes, *collections of agents outperform individuals partially because people see and think about the problems differently.* Additional people create the opportunity for more potential solutions. These additional solutions are only possible if people differ.

[6] See Appendix 1 for a complete description of those job categories we included as problem solvers.

[7] The data for 1950 includes workers age 14 or older and for the other two years includes workers age 16 or older.

If all people encoded and solved problems identically, multiple heads would be no better than one. To say abstractly that diversity can be beneficial seems indisputable, but such statements in no way imply that any particular model of diversity stakes any claim to accuracy. In the case of the present model, the issue of whether it captures those aspects of diversity of thought which are most relevant to group problem solving is left to the reader's discretion.

One way to test whether our approach has merit would be to test whether groups which are more diverse according to our measure actually perform better. Empirical research using experiments with problem solving groups of varying degrees of cultural diversity (Watson, Kumar and Michaelsen 1993) find that groups consisting of more diverse individuals perform better than groups of homogeneous individuals once initial communication barriers have been overcome. If group members value outcomes differently, then diversity may be of little benefit (Chatman, Polzer, Barsade, and Neale 1997). Overall though, there seems to be a strong consensus that diverse groups perform better at problem solving. Robbins (1994) in his organizational behavior textbook says that "the evidence generally supports the conclusion that heterogeneous groups perform more effectively than do those that are homogeneous." Although these studies do not measure diversity in the same way as we do, culture plays a nontrivial role in how we interpret and approach problems. If there exists a positive correlation between cultural diversity as measured by sociologists and problem solving diversity as measured here, then these studies can be viewed as supportive of our second assumption. In a survey article, Thomas and Ely (1996) support this correlation conjecture, saying that "diversity should be understood as *the varied perspectives and approaches to work* that members of different identity groups bring." The *italics* are theirs.

3 A Model of Diverse Problem Solvers

We begin by constructing a model of a finite group of problem solvers of limited ability attempting to maximize a value function defined over a large but finite set of objects (potential solutions to the problem). All problem solvers agree on the value function and they collectively attempt to find an optimal (or satisfactory) solution to the problem. Each problem solver consists of an *internal language*, a *perspective*, and a *heuristic*. A problem solver uses her internal language to represent the objects. Her perspective is a mapping from the objects into her internal language, and her heuristic consists of rules of thumb for moving around the space of objects in her internal language. Throughout this paper, we assume that problem solvers rely on the same internal language – binary strings of length n. Binary strings offer two advantages: they are easy to understand and they map easily into many economic problems. Letting 1 denote "yes" and 0 denote "no", a binary string can denote the set of projects to be undertaken (Page 1996), the group of employees assigned to a task, the attributes of a product, the cities in which a movie is released, or the magazines in which a particular advertisement is going to run.

Formally, the *objects* are binary strings of length n, denoted by $S = \{0,1\}^n$. Each element in a string is referred to as a *bit.* The i-th bit of a string s is denoted by s_i. A value function then maps each object into a real number denoted by $V : S \to \Re$. Every problem solver uses an internal language that is also binary strings of length n. Note that this does not imply that all agents encode objects identically. Agents may differ in how they map objects to strings.

Def'n: *A* **perspective** *M of a problem solver is a 1-1 mapping from the set of objects to the problem solver's internal language. In our binary string model, a* **perspective** *$M : S \to S$ is one–to–one and onto.*

In our definition of a perspective in the binary string model, the set S describes both the domain and the range of the mapping M. As domain, S represents the set of objects. As the range, S represents the objects in the problem solver's internal language. To avoid confusion, we refer to the object 00, for example, as *object string* 00, and if $M(00) = 11$, we refer to 11 as the *M–string* 11. This construction enables the introduction of an identity perspective, denoted by I, which means the problem solver encodes the objects with the identity mapping.

The internal language may be interpreted neurologically or metaphorically. Our brains perceive and store information, and undoubtedly, these perceptions differ across individuals. The alternative, less biological interpretation of an internal language, is that people may interpret problems based on their training, be it as economists, lawyers, etc.. An economist may look at a problem and say "this is a Markov chain."

Agents also have heuristics. A heuristic consists of rules for adapting the status quo solution which might lead to improvement in the problem solver's perception. A problem solver's heuristic might be thought of as her bag of tricks. We want to create a mathematical notion of a heuristic. We build from the concept of an algorithm, which consists of a finite set of instructions (Knuth 1968).

Def'n: *A* **heuristic** *A is a finite collection of mappings, $\{\phi_1, \phi_2, ..., \phi_m\}$, each a mapping from the set S to S, i.e., $A =: \{\phi_1, \phi_2, ..., \phi_m\}$ and for any $k = 1, 2, ..., m$, $\phi_k : S \to S$.*

Given any status quo solution in the agent's perspective, $M(s)$, the agent's heuristic defines a set of solutions $\{\phi_1(M(s)), \phi_2(M(s)), ..., \phi_m(M(s))\}$ where the agent may next find an improvement.

A problem solver is then characterized by $P = (M, A)$ where M is referred to as her perspective and A is called her heuristic. **In this paper, because we rely on a binary string model, the heuristics we describe are rather simple. However, they can be interpreted as more sophisticated problem solving techniques. So, when we have an agent flip a zero to a one, this could be**

interpreted as applying Newton's method or as hill climbing. The only assumptions required are that the set of heurstics an agent could apply does not depend on the current solution and that the heuristics sometimes work and sometimes do not.

Important to the understanding of the paper is the concepts of neighborhood and local optimum which we now define.

Def'n: *The* **neighborhood of an object** *s for $P = (M, A)$,*

$$b_M(A, s) = \{s' \in S : M(s') = \phi(M(s)) \text{ where } \phi \in A\}$$

Def'n: *An object $s \in S$ is a* **local optimum of V with respect to** *$P = (M, A)$, denoted by $s \in L(P, V)$, if and only if*

$$V(s) \geq V(s') \quad \text{for all } s' \in b_M(A, s)$$

In other words, an object is a local optimum for a problem solver if nothing in its neighborhood has a higher value. Note that any maximum of V is a local optimum for every problem solver.

Our general formulation of heuristics needs refining along two dimensions. First, we must clarify how a problem solver applies her heuristic to an object. Second, we must make more explicit what the ϕ's can be. To address the first issue, suppose that the problem solver begins a search from the string $s \in S$ in her internal language by asking, for each ϕ_k, whether $V(M^{-1}(\phi_k(s))) > V(M^{-1}(s))$. Two different search processes come to mind. First, $\phi_j(s)$ becomes the new status quo if and only if it gives the highest value among all the $\phi_k(s)$'s and it has a strictly higher value than s.[8] Second, apply the ϕ_k's in some order to the status quo string s, and let the first $\phi_k(s)$ that has a strictly higher value than s become the new status quo. In each case, the search stops at a solution s in the problem solver's internal language if and only if there is no further improvement from applying the ϕ_k's. Starting from a given initial object, the first and second ways of updating will yield different local optima. Further, the local optima attained from the second updating rule are dependent on the order that the ϕ_k's are applied. However, using either updating rule and regardless of the order chosen, the set of possible outcomes remains unchanged. This set is the set of local optima of the problem solver. **However, the probabilities of landing at these local optima may depend on the problem solving approach. However, with the exception of Section 5 that includes a detailed search process, none of the results depend on the distribution over the local optima. They rely only upon characterizing the set of local optima. Thus, our findings are robust.** For ease of exposition, we use the second updating rule in examples and discussions throughout the paper.

The second issue, what the ϕ_k's can be, is more problematic. Without further restrictions, a ϕ_k could describe a different instruction for each string. In order that

[8] If there is a tie, specific tie-breaking rules will be applied.

our model be tractable, we need a more restrictive notion. We require the rules to be defined independently of the string. Flipping the seventh bit or setting the third bit to a zero would both be rules that are string independent. We will restrict our analysis to heuristics that resemble the former. Without sacrificing simplicity, a rule that flips a bit is more flexible in generating possible solutions than a rule that sets a particular value for a bit. In other words, the former rule encompasses the latter in generating possible solutions. Section 6 aside, none of our results will change if we have included the second type of rules or other string independent rules in addition to the rules that flip bits. Our results are proved through constructions of appropriate problem solvers. Even in Section 6 where we present the idea that explicitly modeling perspectives increases the number of diverse ways of solving a problem, we need only assume that rules are string independent. However, by restricting to rules that flip bits, we are able to present this idea in the most simple and intuitive way.

In our model, in the case of binary strings of length 3, a heuristic might consist of the following three rules which generate a set of neighboring strings given an initial string: flip the first two bits; flip the first and last bits; and flip all three bits. The neighbors of the string 000 generated by this heuristic are 011, 101, and 111.[9] Similarly, the neighbors of the string 001 are 010, 100, and 110. We formalize this idea as follows: a *flipset* is a subset of bits. Mathematically, a flipset maps S onto itself. Let $N = \{1, ..., n\}$.

Def'n: *A* **flipset** $\phi : S \to S$, *where* $\phi \subseteq N$, $\phi(s) = \hat{s}$ *where* $\hat{s}$ *is defined according to the following rule:*

$$\hat{s}_i = \begin{cases} 1 - s_i & \text{if } i \in \phi \\ s_i & \text{if } i \notin \phi \end{cases}$$

Flipsets are an appealing set of rules for both mathematical and intuitive reasons. First, it is straightforward to show that the binary operation of composition in the set of flipsets is both associative and commutative. Second, if we interpret the bits in a string as a collection of attributes, then a single bit flip corresponds to changing an attribute. Similarly, a multiple bit flip represents changing a collection of attributes. If a problem solver believes that bits one and six are related, then she might want to flip both simultaneously.

In this paper, we focus on the class of flipset heuristics which is defined as follows. Let $m \geq 1$.

Def'n: *A* **flipset heuristic** $A = \{\phi_1, \phi_2, ..., \phi_m\}$ *where* $\phi_i \subseteq N$ *is a flipset.*

In this framework, the heuristic which tests to see whether flipping any individual bit would improve can be described by the collection of flipsets, $A_E = \{\phi_1^e, ..., \phi_n^e\}$,

[9] In the binary encoding bits are numbered from right to left. Flipping the first bit of the string 000 gives 001.

where $\phi_i^e = \{i\}$. We refer to this as the *elementary heuristic*. The elementary heuristic can also be defined for subsets of the set of bits: The *elementary heuristic defined on* $K \subseteq N$ is defined by $A_E^K = \{\phi_i^e : i \in K\}$. This becomes important when a heuristic does not apply to the entire domain, such as when a large problem is broken down into subproblems.

The number of flipsets in a flipset heuristic determines the size of the search neighborhood. Assume that all flipsets in a flipset heuristic are unique.

Def'n: *The* **size** *of a flipset heuristic* $sz(A) = \mid A \mid$

Note that the elementary heuristic is of size n. From this point on in the paper, all heuristics are assumed to be flipset heuristics.

We assume that each problem solver has her own perspective/heuristic pair that captures her expertise in locating a good solution. With this assumption, a collection of agents have an advantage over individuals in reaching a better solution simply because different perspective/heuristic pairs lead to the examination of more potential solutions and thus a better final solution. We next present a simple example to illustrate our basic model of problem solving and the intuitive idea of why a group of heterogeneous agents can locate a better solution than individual agents separately.

Example: Three public projects, $\{p_1, p_2, p_3\}$, are under consideration by a team of city council members. The issue is which, if any, of the projects to fund. There are eight potential solutions. The net value for the city of each solution is provided in the following table. The goal of the team is to locate the solution that generates the highest value for the city.

objects (possible solutions)	values
fund none	0
fund p_1 only	40
fund p_2 only	20
fund p_1 and p_2	60
fund p_3 only	30
fund p_1 and p_3	50
fund p_2 and p_3	70
fund all three projects	10

Suppose council member 1 has the following perspective/heuristic pair (M_1, A_1): She encodes possible solutions into binary strings of length 3, $s = s_3s_2s_1$, where for any $i = 1, 2, 3$, $s_i = 1$ if and only if p_i is funded. Her perspective M_1 is summarized in the table below. Her heuristic A_1 is the elementary heuristic $A_1 = \{\{1\}, \{2\}, \{3\}\}$. Applying flipset $\{1\}$ changes the decision about p_1 only.

objects	$M_1(\cdot)$
fund none	000
fund p_1 only	001
fund p_2 only	010
fund p_1 and p_2	011
fund p_3 only	100
fund p_1 and p_3	101
fund p_2 and p_3	110
fund all three projects	111

Let's look at a possible search path of Member 1. Suppose she starts at "fund no projects" which she encodes into 000 and its value equals 0. She updates the status quo string by applying her heuristic sequentially, say in the order of $\{1\}, \{2\}$, and $\{3\}$. Applying $\{1\}$ to the string 000 will lead her to 001 which corresponds to "fund p_1 only" and it has a value 40 which is an improvement. So she updates the status quo string to 001. She then applies $\{2\}$ to the string 001 leading to 011 which corresponds to "fund p_1 and p_2" with value 60. She then updates the status quo to 011. Applying $\{3\}$ to 011 leads to 111 which corresponds to "fund all projects" and has a value 10. No updating occurs this time. Applying $\{1\}$ or $\{2\}$ to 011 will find no improvement either, so the search ends. The solution of her search is "fund p_1 and p_2" with value 60. The object "fund p_1 and p_2" is a local optimum for Member 1. It is easily seen that the set of local optima for Member 1 with (M_1, A_1) is {fund p_1 and p_2, fund p_1 and p_3, fund p_2 and p_3} in which "fund p_2 and p_3" is the maximum of the value function. Member 1 with her perspective/heuristic pair will always end up with one of the three local optima. The initial object and the order that her flipsets are applied determine which local optimum will be reached.

Suppose Member 2 has a different perspective/heuristic pair (M_2, A_2): She also encodes possible solutions into binary strings of length 3, $s = s_3s_2s_1$, but for her, $s_i = 1$ if and only if p_i is *not* funded. M_2 can be summarized into a table (we omit that). A_2 consists of two flipsets, $A_2 = \{\{1,2\},\{1,3\}\}$. Applying $\{1,2\}$ will change the decisions on p_1 and p_2 simultaneously. $\{1,3\}$ has a similar interpretation.

It is easily seen that the set of local optima for Member 2 with (M_2, A_2) is {fund p_1 only, fund p_2 and p_3} which differs from Member 1's set of local optima. The maximum of the value function, "fund p_2 and p_3", is in both sets. It is a general phenomenon that the maximum of a value function is contained in every problem solver's set of local maxima.

If Member 1 and Member 2 work separately, either may end up with a solution that is not the best solution. Imagine they work collectively as follows: Member 1 works on the problem first and locates a solution; Member 2 then joins Member 1 and they work together as one person. Without too much difficulty, one can see that by working together Member 1 and Member 2 locate the best solution, "fund p_2 and p_3". For example, suppose Member 1 works on the problem and gets stuck with "fund p_1 and p_3" which is one of her local optima. Member 2 joins her and takes the solution located by Member 1. In Member 2's encoding, "fund p_1 and p_3" is 010 string. Applying flipset $\{1, 2\}$, Member 2 reaches string 001 which corresponds to "fund p_2 and p_3" with value 70 and is the best solution. Because of the different perspective/heuristic pair that Member 2 has, she is able to improve upon the solution where Member 1 gets stuck.

The above example, though simple and artificial, illustrates the basic idea of how differences in perspective/heuristic pairs can improve performance. A collection of problem solvers only become stuck at a point if no member can locate a higher valued object. This means that groups only get stuck at an object if every member would be stuck there as well. Owing to path dependency, evaluating the expected performance of an individual or a group requires a description of how an individual problem solver applies her heuristic and how a group of problem solvers apply their heuristics to problems. These descriptions are provided later in the paper. One way to envision the search process is as a Markov chain. Through their perspectives and heuristics, a problem solver applied to a problem creates a transition matrix: a probability of going from any object to any other object. Together with a prior probability distribution over objects, the Markov process generates a probability distribution over the set of local optima allowing for the computation of the expected value of a local optimum for a problem solver. The expected value of a group of problem solvers can be calculated similarly.

Of course, agents with distinct perspective/heuristic pairs do not necessarily locate different solutions. We can construct two problem solvers with distinct perspective/heuristic pairs creating identical neighborhoods and thus the same set of local optima. In such a case, these two problem solvers will not be of any help to each other in locating better solutions even though they differ.

We shall say that two problem solvers are *equivalent* if their perspectives and heuristics generate identical neighborhood structure on the set of objects.

Def'n: *Let $P_1 = (M_1, A_1)$ and $P_2 = (M_2, A_2)$. P_1 and P_2 are* **equivalent** *if for any $s \in S$, $b_{M_1}(A_1, s) = b_{M_2}(A_2, s)$.*

In section 6, we discuss in greater detail the issue of the equivalence of problem solvers. If two problem solvers create distinct neighborhood structures, they will solve problems differently. Thus, we measure the diversity of two problem solvers by the average number of objects that belong to only one person's neighborhood, and not by the perspectives and heuristics directly.

Def'n: *The* **diversity** *of $P_1 = (M_1, A_1)$ and $P_2 = (M_2, A_2)$,*

$$\Delta(P_1, P_2) = \sum_{s=1}^{2^n} \frac{|(b_{M_1}(A_1, s) \cup b_{M_2}(A_2, s)) \backslash (b_{M_1}(A_1, s) \cap b_{M_2}(A_2, s))|}{2^n}$$

Remark 1 *Two problem solvers P_1 and P_2 are equivalent iff $\Delta(P_1, P_2) = 0$.*

Def'n: *A group of problem solvers, $\{P_1, ..., P_a\}$ where $P_k = (M_k, A_k)$ for any $k \in \{1, ..., a\}$, is* **maximally diverse** *if $\Delta(P_i, P_j) = |A_i| + |A_j|$ for all i and j in $\{1, ..., a\}$ such that $i \neq j$.*

We illustrate these concepts by elaborating on our earlier example. The two city council members are denoted by $P_1 = (M_1, A_1)$ and $P_2 = (M_2, A_2)$. M_1 and M_2 are described below, $A_1 = \{\{1\}, \{2\}, \{3\}\}$ and $A_2 = \{\{1, 2\}, \{1, 3\}\}$. For notational convenience, we identify the binary strings by integers using the standard mapping. In what follows, s denotes a binary string as well as its corresponding integer.

s	0	1	2	3	4	5	6	7
s	000	001	010	011	100	101	110	111

$\cup_s \backslash \cap_s$ denotes $|(b_{M_1}(A_1, s) \cup b_{M_2}(A_2, s)) \backslash (b_{M_1}(A_1, s) \cap b_{M_2}(A_2, s))|$ in the table below. We conclude that these two problem solvers are maximally diverse.

s	$M_1(s)$	$b_{M_1}(A_1, s)$	$M_2(s)$	$b_{M_2}(A_2, s)$	$\cup_s \backslash \cap_s$
0 : fund none	0	1,2,4,	7	3,5	5
1 : fund p_1 only	1	0,3,5	6	2,4	5
2 : fund p_2 only	2	0,3,6	5	1,7	5
3 : fund p_1 and p_2	3	1,2,7	4	0,6	5
4 : fund p_3 only	4	0,5,6	3	7,1	5
5 : fund p_1 and p_3	5	1,4,7	2	0,6	5
6 : fund p_2 and p_3	6	2,4,7	1	5,3	5
7 : fund all three projects	7	3,5,6	0	4,2	5
$\Delta(P_1, P_2)$					5

4 Optimality Through Diversity

In this section, we discuss our results on diversity and optimality. The two propositions below demonstrate how diversity among individual agents of limited ability can lead to optimal collective solutions. Here we must be careful as to what we mean by a collective and what we mean by a solution. There are two possible interpretations along each dimension. A solution could be a tangible output that can be sold in a market such as a better mousetrap, or it could be an improvement in a product such as a movie or a pharmaceutical product, that is not yet at the market stage. A collective can be the entire economy or it can be a small group working independently.

Absent incentive problems our model applies equally well to each of these interpretations. However, the inclusion of incentives makes some interpretations more compelling than others. Tangible product problems being solved by the entire economy create few incentive problems other than strategic delay. If someone discovers a better cure for the common cold, she does so and takes her rents. Compare this situation to that of a team member working for a software company . If the team member sees how to improve the software, he too should have an incentive to reveal his solution. The difference lies in the strength of the incentive. In the former case, if the person can not better treat the cold, then she earns no rents. In contrast, the team member may receive the same pay regardless of whether she offers the improvement. So, the latter case allows the possibility of free riding. We assume, for the purposes of this paper which focus on diversity in problem solving, that the firm can structure incentives so that team members reveal improvements.

We dichotomize the notion of diversity along the perspective and heuristic dimensions. Our first proposition states that there exists a perspective such that the elementary heuristic has a unique local optimum, namely the global optimum. Thus, for any problem there exists a way of viewing the problem, a perspective, such that the simple rule of thumb of flipping individual bits locates the optimal solution.

Proposition 1 *For any function V such that $V(s) \neq V(s')$ for any $s \neq s'$, there exits an M such that $\mid L((M, A_E), V) \mid = 1$. (Recall from Section 3 that $L((M, A_E), V)$ denotes the set of local optima of the problem solver (M, A_E).)*

Pf: We can order the strings according to their values under V from s^1 to s^{2^n} where $V(s^j) > V(s^{j+1})$. We next construct the linear function V_L as follows:

$$V_L(s) = \sum_{i=1}^{n} 2^{i-1} \cdot s_i$$

We can also order the strings according to their value under V_L from $\hat{s}^1$ to $\hat{s}^{2^n}$ where $V_L(\hat{s}^j) > V_L(\hat{s}^{j+1})$. Define the perspective M as follows: $M(s^j) = \hat{s}^j$ for j equal 1 to 2^n. Choose an object string $s \in L((M, A_E), V)$. It suffices to show that $M(s)_i = 1$ for i equal 1 to n. The proof proceeds by contradiction. Suppose that there exists an i such that $M(s)_i = 0$. By assumption, $V_L(\phi_i^e(M(s))) > V_L(M(s))$. It follows that $V(M^{-1}(\phi_i^e(M(s))) > V(s)$, which contradicts $s \in L((M, A_E), V)$.

Proposition 1 suggests the possibility that an employee can locate an optimal solution to a difficult problem even though she uses an unsophisticated heuristic, so long as she has the right perspective. A second implication is that a diversity of perspectives can lead to the location of the global optimum. And a third implication concerns the categorization of the difficulty of a function. Since every function is easy for someone, function difficulty can only be measured relative to the perspective of the problem solver.

We next consider the case where perspective is fixed but heuristics can vary. We state an observation and then a proposition. The observation is that diversity in heuristics guarantees optimality. The proposition refines this observation by placing a bound on the number of heuristics as a function of the difficulty of the problem.

Observation: *Given a value function V and a common perspective M, there exists a set of problem solvers possessing different heuristics which locate the optimal solution irrespective of the starting point.*

To see why this observation must be true, let each agent's set of heuristic consist of a single flipset. Assume that there are $2^n - 1$ agents and each possesses a unique heuristic. Collectively, they are guaranteed to locate the optimum. This proof relies on exhaustive search, making it somewhat unsatisfying. Ideally, we would like a relationship between the number of heuristics needed. We can prove a stronger result which borrows results from Page (1994,1996). He constructs two measures of difficulty for functions defined over binary strings based upon cover theory. The first of these measures, *cover size,* captures the difficulty of solving a problem in parallel. The second measure, *ascent size,* captures the difficulty of solving a problem using a hill climbing (or ascent) algorithm. The second of these measures is more appropriate for the analysis considered here. Essentially, a problem has an ascent size of one for a perspective if and only if the elementary heuristic has a unique local optimum which is the global optimum. It has an ascent size of two if and only if the global optimum is located as long as the combined set of heuristics contains all individual bit flips and all pairs of bit flips. An ascent size of k means that a collection of agents whose flipsets contain all sets of k bits or less will locate the optimal solution.

A problem is considered easy if the ascent size equals one and difficult as the ascent size approaches n, the number of bits.[10] Thus, ascent size can be interpreted as the appropriateness of a perspective. The smaller the ascent size, the better the perspective for the function.

Proposition 2 *If a function V has an ascent size of k given a perspective M, then an upper bound on the minimal number of problem solvers with m flipsets and the common perspective M is given by*

$$b_k = \lfloor \frac{\sum_{j=1}^{k} \frac{n!}{(n-j)!j!}}{m} \rfloor$$

where $\lfloor x \rfloor$ is the least integer greater than x.

Pf: By assumption the problem has an ascent size of k. Therefore, the only local optimum relative to flips of k bits or less is the global optimum. The number of flipsets of size k or less equals:

[10] Ascent size can also be measured relative to upper contour sets. Page (1996) shows that as the function value improves, the ascent size weakly decreases.

$$\sum_{j=1}^{k} \frac{n!}{(n-j)!j!}$$

Therefore, b_k gives an upper bound on the minimal number of problem solvers needed.

The reason that b_k is an upper bound is that not all of the flipsets may be necessary. Some subset of flipsets may be sufficient to locate the optimum. An implication of the proposition is that as problems become harder more problems solvers are needed to guarantee locating the optimum. In addition, these problem solvers must be diverse. If they are not maximally diverse, then even more will be needed to locate the optimum.

5 Problem Solving Firms

In Section 2, we presented data showing an increase in the proportion of workers who solve problems. A natural question to ask is whether constant, decreasing, or increasing marginal returns would hold for problem solving firms. Although casual intuition suggests that problem solving would yield decreasing returns to additional problem solvers, several insights emerge from our analysis which lead us to question such a conclusion. Ignoring coordination and communication problems, we first find that *any* "production function" is possible. By this we mean, given any nondecreasing function f defined over a finite set of integers, we can construct a value function and a group of problem solvers that generate f when sequentially applied, even when the workers are of "equal ability." Two individuals having equal ability means that they have identical expected values of search. Second, we can construct value functions and problem solvers of equal ability who generate both increasing and decreasing returns depending upon the order that they are applied to the problem.

These results establish the possibility of these phenomena but do not address their likelihood. In this way, they are similar to the findings of Sonnenschein (1973), Mantel (1974), and Debreu (1974) that show the possibility of any aggregate excess demand function. These demand function findings rely on manipulating income effects. We apply a similar trick on the basins of attraction of the local optima of various problem solvers. Our results should not be interpreted as negative. Instead, they should be seen as suggestive that problem solving has less regularity than manufacturing: problem solving does not fall into a particular class of marginal returns. While problems of coordination and communication in group problem solving may be amenable to formal analysis using our framework, by analyzing marginal returns in the absence of these costs, we point to the behavior of an upper bound of marginal returns.

We also explore the types of assumptions required to generate diminishing marginal returns for every sequence of a given group of problem solvers. We find that one set of sufficient conditions – that a priori none of the problem solvers has any insight into the problem at hand – runs counter to basic economic intuition. Successful problem solving firms confront problems for which their employees have some expertise. Within the context of our model this expertise takes the form of perspective/heuristic pairs for which the set of local optima is small and of high average value.

Proofs of the results of this section are all contained in Appendix 2.

5.1 Group Problem Solving

Problem solving firms attempt to locate high valued solutions. Given a value function, we measure the output of a problem solving firm consisting of a group of problem solvers by the expected value of the local optima for the group. This performance measure depends on the way we model how an individual searches for a solution and how a group of agents work together to search for a solution. We assume that individual problem solvers apply their flipsets sequentially. Any time a problem solver obtains an improvement, she updates the status quo.

Specifically, given a value function V and a problem solver $P = (M, A)$ where A is a heuristic consisting of m flipsets $\{\phi_1, \phi_2, ..., \phi_m\}$, the outcome of the problem solver's search depends on her starting point and the order in which she applies her flipsets.[11] Let σ be a permutation on $\{1, 2, ..., m\}$ denoting that flipsets are being applied in the order of $\phi_{\sigma(1)}, \phi_{\sigma(2)}, ..., \phi_{\sigma(m)}$. Let $A^\sigma(s, M)$ denote the outcome of the search where the search starts at the object string s and flipsets are applied in the order σ. The outcome $A^\sigma(s, M)$ is reached in the following way: First, the problem solver encodes all objects including the starting point s into her internal language according to M; Then in her encoded problem space, she sequentially applies her flipsets in the order $\phi_{\sigma(1)}, \phi_{\sigma(2)}, ..., \phi_{\sigma(m)}$. She continues cycling through her flipsets in that order until none of them locate a higher valued string, at which point the search stops. Finally, $A^\sigma(s, M)$ is the solution in the original problem space corresponding to the point where the search stops. Formally,

Def'n: *Let $\sigma \in \Sigma(\{1, 2, ..., m\})$, where $\Sigma(\{1, 2, ..., m\})$ is the permutation group of the set $\{1, 2, ..., m\}$. The* **outcome of the search** *by problem solver $P = (M, A)$, where $A = \{\phi_1, \phi_2, ..., \phi_m\}$, applying flipsets to an initial object string s in the order of $\phi_{\sigma(1)}, \phi_{\sigma(2)}, ..., \phi_{\sigma(m)}$, denoted by $A^\sigma(s, M)$, is defined as follows:*

Step 1: $a = 0$, $t = 0$, $s' = M(s)$

Step 2: $t = t + 1$, $s^t = s'$

Step 3: If $t > m$ and $s^{t-m} = s^t$ then go to Step 6 else go to Step 4

Step 4: Let $a = a + 1$. If $a > m$ then let $a = 1$

[11] Alternatively, we can have problem solvers with their orders of filpsets fixed. It does not change our results. In Hong and Page (1998), the order of flipsets of a problem solver is fixed.

Step 5: If $V(M^{-1}(\phi_{\sigma(a)}(s^t))) > V(M^{-1}(s^t))$ *then* $s' = \phi_{\sigma(a)}(s^t)$
else $s' = s^t$. *Go to Step 2*

Step 6: $A^\sigma(s, M) = M^{-1}(s')$. *End.*

Recall from Section 3 that the set of local optima of a problem solver P for a value function V, $L(P, V)$, is the set of solutions from which the problem solver can no longer find a higher valued solution by applying any of her flipsets. Clearly, for any starting point s and any order σ, $A^\sigma(s, M) \in L(P, V)$. A different starting point and/or a different order of flipsets may lead to a different local optimum, but a search will end if and only if it hits a local optimum. We measure performance of an individual problem solver by the expected value of these stopping points assuming a uniform distribution over starting points and a uniform distribution over the permutation group (i.e. assuming each object is equally likely to be a starting point of a search and each order is equally likely to be the order chosen). In other words, performance of a problem solver is measured by the expected value of her local optima where the probability distribution of the local optima is induced by the uniform probability distributions over starting points and orders.

Def'n: *The* **expected value of local optima** *for problem solver* $P = (M, A)$, *where* $A = \{\phi_1, \phi_2, ..., \phi_m\}$, *is*

$$E[V : P] = E[V(A^\sigma(s, M)) : s \in S \text{ and } \sigma \in \Sigma(\{1, 2, ..., m\})]$$

where s *and* σ *are drawn independently according to uniform distributions on* S *and* $\Sigma(\{1, 2, ..., m\})$ *respectively.*

When a group of problem solvers work together, they could apply their heuristics sequentially or simultaneously. Either results in path dependence of outcomes except in special cases such as unique local optima. Though path dependence implies that different rules may yield different expected values, the rule chosen does not appear to be that important for our results. More formally, the propositions below hold for several alternative and equally plausible descriptions of group behavior. The constructive nature of the proofs suggests that our results might hold for *any* description of group problem solving. For mathematical convenience, we assume that the group of problem solvers act sequentially. The first problem solver applies her heuristic until attaining a local optimum. The second problem solver then joins the first problem solver. Together, the two act as a single problem solver with a combined set of flipsets. When that subgroup has located a local optimum, the next problem solver joins, and the three act as a single problem solver.

This formulation of group problem solving relies crucially on the idea of sequentially applying a combined set of flipsets, each possibly originating from a different problem solver. Recall that a flipset of an agent is an operation defined for that particular agent's encoded problem space. If all agents in the group have the same encoding or perspective M, then applying a set of flipsets sequentially even when they originate from different agents poses no problem – the notion is just a trivial extension of $A^{\sigma}(s, M)$. However, when agents have different perspectives, we need to keep in mind that flipsets should be applied to strings in their respective perspectives. A natural way to keep track of the search progress when flipsets are defined for different perspectives is to always specify the status quo string in the original problem space, common to every agent. We give a formal definition. In the following definition, each flipset is identified by two subscripts, the first refers to the agent, the second refers to the flipset in the agent's heuristic.

Let $\{P_j\}_{j=1}^{i}$ where $i > 1$ be a group of problem solvers. For each problem solver $j \in \{1, ..., i\}$, $P_j = (M_j, A_j)$where $A_j = \{\phi_{j1}, \phi_{j2}, ..., \phi_{jm_j}\}$. Consider a combined set of flipsets, $\{\phi_{h_1 g_{h_1}}, \phi_{h_2 g_{h_2}}, ..., \phi_{h_l g_{h_l}}\}$. For any $1 \leq a \leq l$, $\phi_{h_a g_{h_a}}$ denotes problem solver h_a's flipset number g_{h_a}. Obviously for any $1 \leq a \leq l$, $h_a \in \{1, ..., i\}$ and $g_{h_a} \in \{1, ..., m_{h_a}\}$.

Def'n: *The* **outcome of a search** *by problem solvers* $\{P_j\}_{j=1}^{i}$ *applying their flipsets* $\{\phi_{h_1 g_{h_1}}, \phi_{h_2 g_{h_2}}, ..., \phi_{h_l g_{h_l}}\}$ *to an initial object string* s *in that exact order is defined by the following steps:*

Step 1: $a = 0,\ t = 0,\ s' = s$

Step 2: $t = t + 1,\ s^t = s'$

Step 3: If $t > l$ *and* $s^{t-l} = s^t$ *then go to Step 6 else go to Step 4*

Step 4: $a = a + 1$ *if* $a > l$ *then* $a = 1$

Step 5: If $V(M_{h_a}^{-1}(\phi_{h_a g_{h_a}}(M_{h_a}(s^t)))) > V(s^t)$,
then $s' = M_{h_a}^{-1}(\phi_{h_a g_{h_a}}(M_{h_a}(s^t)))$ *else* $s' = s^t$. *Go to Step 2*

Step 6: Outcome $= s'$. *End.*

The notation is messy here but the idea is similar to the definition of $A^{\sigma}(s, M)$. One only needs to note that the flipsets under consideration come from different agents and therefore are applied to strings in their respective perspectives.

We next define performance of a group of problem solvers by the expected value of local optima of the group. In the definition, we introduce the following notation: for any $1 \leq k \leq i$, $\sigma_k \in \Sigma(I_k)$ where $\Sigma(I_k)$ denotes the permutation group on I_k and $I_k = \cup_{j=1}^{k}\{j1, ..., jm_j\}$, and $U_k^{\sigma_k}(s)$ denotes the outcome of the search by problem solvers applying $\{\phi_{\sigma_k(11)}, ..., \phi_{\sigma_k(1m_1)}, ..., \phi_{\sigma_k(k1)}, ..., \phi_{\sigma_k(km_k)}\}$ to s in that order. Here, I_k is the index set of all flipsets of the first k agents, σ_k is a permutation of I_k which specifies the order in which flipsets of the first k agents are applied. By definition, when $k = 1$, $U_1^{\sigma_1}(s)$ is the outcome of the search by problem solver 1 starting at s and applying her flipsets in the order σ_1, i.e., $U_1^{\sigma_1}(s) = A^{\sigma_1}(s, M_1)$. Therefore, $U_1^{\sigma_1}(s) \in L(P_1, V)$. Let's now consider $U_2^{\sigma_2}(U_1^{\sigma_1}(s))$. $U_2^{\sigma_2}(U_1^{\sigma_1}(s))$ is the outcome of the following search process by agents 1 and 2: Agent 1 starts her search at s and applying her flipsets in the order of σ_1 and gets stuck at $U_1^{\sigma_1}(s)$; agent 2 then joins agent 1 and the two agents combine their flipsets and applying them in the order of σ_2 with $U_1^{\sigma_1}(s)$ (the outcome of the search by agent 1 alone) as their starting point. This search will stop at an object from which no flipset from the combined set of flipsets of the two agents will find a higher valued object. Therefore $U_2^{\sigma_2}(U_1^{\sigma_1}(s))$ is a local optimum for both agent 1 and agent 2. That is, $U_2^{\sigma_2}(U_1^{\sigma_1}(s)) \in L(P_1, V) \cap L(P_2, V)$. Finally, $U_i^{\sigma_i}(...U_2^{\sigma_2}(U_1^{\sigma_1}(s))...)$ is similarly defined to be the outcome of the search by the group of problem solvers $\{P_j\}_{j=1}^i$ when problem solvers are applied sequentially in the order of 1 to i. We have, $U_i^{\sigma_i}(...U_2^{\sigma_2}(U_1^{\sigma_1}(s))...) \in \cap_{j=1}^i L(P_j, V)$. The set $\cap_{j=1}^i L(P_j, V)$ is called the set of local optima of the group $\{P_j\}_{j=1}^i$. As earlier, we assume that each object is equally likely to be the starting point of a search and in various stages of group search, each order is equally likely to be the chosen order. We define performance of the group $\{P_j\}_{j=1}^i$ by the expected value of local optima of the group.

Def'n: *The* **expected value of local optima** *for the group of problem solvers* $\{P_j\}_{j=1}^i$ *where* $i > 1$

$$E[V : \{P_j\}_{j=1}^i] = E[V(U_i^{\sigma_i}(...U_2^{\sigma_2}(U_1^{\sigma_1}(s))...) : s \in S \text{ and } \sigma_k \in \Sigma(I_k) \text{ for } 1 \leq k \leq i]$$

where s and σ_k's are drawn independently according to uniform distributions on S and $\Sigma(I_k)$'s respectively.

By defining group problem solving by adding members sequentially, we can compute the marginal value of adding problem solvers. We call this the marginal product of problem solvers (MPPS). The MPPS for problem solver number four is his contribution to the expected value when added to the first three problem solvers. Given our assumption that problem solvers differ, no general MPPS exists per se. Each problem solver creates her own increase to total product which depends upon the other problem solvers hired previously. We would have to write MPPS(i, J), where i is the problem solver, and J is an ordered set of problem solvers previously hired. To keep notation at a minimum, we shall speak of returns to adding problem solvers when discussing MPPS(i, J).

5.2 Returns to Additional Problem Solvers

Our next proposition states that we can find a value function and a group of employees such that if the employees are hired in one order the returns to additional problem solvers are decreasing, and if they are hired in another order, the returns to additional problem solvers are increasing. This would not be a surprising result if the problem solvers differed in their abilities. Hiring smarter (dumber) workers first would create decreasing (increasing) returns. However, in our result, all problem solvers have equal ability as measured by their expected value of local optima.

There are several ways to cast this counter-intuitive result. We first provide an example where order matters. In this example, each new worker can escape one of the local optima of the previous group. The order determines where the former local optima go. In one order, the local optima lie in the basin of attraction of the optimal object which has a drastically higher value. Therefore, the returns to additional workers are linear. In another order the local optima lie in the basins of attraction of local optima with only slightly higher values. In this case, the returns to additional problem solvers are small until the last problem solver is added. The return to the last problem solver is very large, so the returns increase.

Example: $n = 6$, the number of problem solvers $m = 22$. All problem solvers use the elementary heuristic but they have different perspectives. All but 23 strings have values equal to zero. We denote these by x_i for $i = 0$ to 22. The value of x_0 is 1 and the value of string x_i is $i \cdot \epsilon$ for $i = 1$ to 22 where ϵ is a small but positive number. The strings with exactly three ones, of which there are twenty, play a prominent role in this example. Let s^3_k where $k = 2$ to 21 denote these twenty strings.

We write problem solver 22's perspective as follows:

$$M_{22}(x_0) = 111111$$
$$M_{22}(x_1) = 000000$$
$$M_{22}(x_k) = s^3_k \text{ for } k = 2 \text{ to } 21$$
$$M_{22}(x_{22}) = 111110$$

Problem solver j for $j = 1$ to 21 has the perspective:

$$M_j(x_0) = 000000$$
$$M_j(x_j) = 111110$$
$$M_j(x_{j+1}) = 111111$$
$$M_j(x_{\lfloor j+k \rfloor}) = s^3_k \text{ for } k = 2 \text{ to } 21 \text{ where } \lfloor j+k \rfloor = (j+k) \, mod \; 22.$$

We can construct the following probability table for reaching the x_i's

Probability of Local Optimum (over 640) a = either 24 or 25 below											
PS	x_1	x_2	x_3	x_4	x_5	x_{10}	x_{11}	x_{20}	x_{21}	x_{22}	x_0
dir	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	sink
22	70	a	a	a	a	a	a	a	a	0	80
21	a	a	a	a	a	a	a	a	0	80	70
20	a	a	a	a	a	a	a	0	80	a	70
.	.	.	.	.	.	.	.	.	.	.	.
3	a	a	0	80	a	a	a	a	a	a	70
2	a	0	80	a	a	a	a	a	a	a	70
1	0	80	a	a	a	a	a	a	a	a	70

Consider the order P_1, P_2. There is only a small increase until problem solver 22 is added at which point there is a huge increase. Therefore, the returns to additional workers are increasing. If the order is changed to P_{22}, P_{21} then there is a large "linear" increase with the addition of each worker. The returns to additional workers is approximately constant.

We can now state the general proposition, which relies on a constructive proof.

Proposition 3 *For any $m \geq 3$, there exist an $n \geq 1$, a value function V defined on binary strings of length n, a group of problem solvers $\{P_i\}_{i=1}^m$, and σ, σ', two elements of the permutation group on $\{1, ..., m\}$, such that (i), (ii), and (iii) hold at the same time:*

(*i*) $E[V : \{P_{\sigma(j)}\}_{j=1}^{i}] - E[V : \{P_{\sigma(j)}\}_{j=1}^{i-1}] \geq E[V : \{P_{\sigma(j)}\}_{j=1}^{i-1}] - E[V : \{P_{\sigma(j)}\}_{j=1}^{i-2}]$ *for* $i = 3, ..., m$

(*ii*) $E[V : \{P_{\sigma'(j)}\}_{j=1}^{i}] - E[V : \{P_{\sigma'(j)}\}_{j=1}^{i-1}] \leq E[V : \{P_{\sigma'(j)}\}_{j=1}^{i-1}] - E[V : \{P_{\sigma'(j)}\}_{j=1}^{i-2}]$ *for* $i = 3, ..., m$

(*iii*) $E[V : P_i] = E[V : P_j]$ *for all* $i, j = 1, ..., m$

If we consider a single problem solver with multiple, equally effective algorithms, then we can reinterpret this result to mean that the returns from an additional algorithm need not be diminishing. The next proposition states that for any non-decreasing function defined over the first m integers there exist a value function, a set of problem solvers of equal ability and an order which generates the function. This implies that arbitrary returns to adding problem solvers are possible even if the problem solvers have equal ability.

Proposition 4 *For any $m \geq 2$ and any function $f : \{1, 2, ..., m\} \rightarrow Q$ (Q denotes the set of rational numbers) which is weakly increasing with $f(1) = 0$ and $f(m) = 1$, there exist an $n \geq 1$, a value function $V : \{0,1\}^n \rightarrow \Re$, and problem solvers $\{P_i\}_{i=1}^m$ such that $E[V : \{P_i\}_{i=1}^j] = f(j)$ for $j = 1$ to m and that $E[V : P_i] = E[V : P_j]$ for all i and j.*

Taken together these two propositions demonstrate the arbitrariness in marginal products of labor in problem solving, but they do not indicate the likelihood of such phenomena, nor do they describe what happens on average. Herein lies a difficulty. If we consider "on average" to be the expected performance of agents solving a value function that is randomly drawn from the space of all functions using a uniform prior, then it is as if the agents are facing a random function – a function defined over a discrete domain taking on random values. Such random functions are *maximally difficult.*

5.3 Maximally Difficult Functions

A *maximally difficult* function has random values from the problem solver's perspective (Macken, Hagan and Perelson 1990). If a function is maximally difficult for a problem solver, then the problem solver has no insight or understanding of the problem.

Proposition 5 *If a function V, taking values in [0,1], is maximally difficult for each problem solver among a group of m problem solvers with identical perspectives, identical number of flipsets in each agent's heuristic and that are maximally diverse, the expected value of local optima of the group is $1 - \frac{\theta}{m}$, where θ is some constant. (See Section 3 for the definition of a group of problem solvers being maximally diverse.)*

For a proof of this proposition, see Macken, Hagan and Perelson (1990). Clearly then, with maximally difficult functions, problem solving with a group of maximally diverse problem solvers will have decreasing marginal returns. It also follows from the proof of this proposition that a group of maximally diverse problem solvers do better on a maximally difficult problem than another group with the same feature except that problem solvers are not maximally diverse. This states that *on average* maximally diverse groups perform better. Of course, it does not imply that for a given problem they necessarily do better. Following is an example illustrating this point.

Example: Consider the following value function $V : \{0, 1, 2, 3, 4, 5, 6, 7\} \rightarrow \Re$:

$V(0)$	$V(1)$	$V(2)$	$V(3)$	$V(4)$	$V(5)$	$V(6)$	$V(7)$
20	60	80	50	40	30	70	90

Suppose that there are two groups of agents trying to locate the maximum of the value function. Let group a consist of P_1 and P_2 and group b consist of P_3 and P_4, where $P_1 = (I, A_1)$, $P_2 = (I, A_2)$, $P_3 = (I, A_E)$, and $P_4 = (I, A_4)$. Let $A_1 = \{\{1,2,3\},\{1,2\},\{1,3\}\}$, $A_2 = \{\{3\},\{1,2\},\{2,3\}\}$, and $A_4 = \{\{1,2,3\},\{1,2\},\{2,3\}\}$.

Given these perspectives and heuristics of the agents, it can be verified that (1) $\triangle(P_1, P_2) = 4$ and $\triangle(P_3, P_4) = 6$, so that group b is maximally diverse but group a is not; (2) Group a (P_1 and P_2 together) can always locate the global optimum; (3) Group b (P_3 and P_4 together) cannot always locate the global optimum, for example, the object string 2 (which has a value of 80) is a local optimum for both P_3 and P_4. Therefore, $E[V : P_3, P_4] < E[V : P_1, P_2] = 90$.

To summarize, if, on average, agents have no insights into problems they try to solve, their average performance can be represented by their performance on solving a maximally difficult function. This makes little economic sense. People sort into professions according to their abilities. Thus, problem solvers should have some insight, intuition, or experience which enables them to perceive problems and choose heuristics which outperform random draws. Therefore, the problems they choose should not be maximally difficult for them. Moreover, the problems should not be easy either. Problem solvers should not be so intelligent as to be able to locate optimal solutions to problems single handedly.

6 Equivalence

Though we differentiate problem solvers along two dimensions, we do not rule out the possibility of two problem solvers with different perspectives and heuristics generating the same neighborhood structure. In this section, we formally define an equivalence relation on perspective/heuristic pairs and derive three propositions that establish the following two facts: First, in the binary string model, allowing different perspectives contributes nontrivially to enlarge the set of possible ways of solving a problem; this further highlights the significance of explicitly allowing for different perspectives. Second, we find an enormous number of equivalence classes. This second observation legitimizes the premise of this paper, that individuals solve problems differently.

The results require substantial mathematical preliminaries. Here we introduce a minimal amount of notation so that we can state and interpret our results. We leave the detailed discussion for Appendix 2. Recall that n is the length of binary strings in our model, and m is the size of a flipset heuristic ($A = \{\phi_1, ..., \phi_m\}$). We place no other restrictions except that $m \leq 2^n - 1$ (this has to hold because we assumed implicitly that all flipsets in A are different). Here, we consider the case where $m \leq n$ so that we can use the elementary heuristic as our benchmark.

Let $1 \leq m \leq n$. Given a flipset heuristic $A = \{\phi_1, ..., \phi_m\}$, and K a subset of $\{1, 2, ..., m\}$, we can define the flips within K.

Def'n: *Given a heuristic* $A = \{\phi_1, ..., \phi_m\}$ *and* $K = \{i_1, i_2, ..., i_k\}$ *a subset of* $\{1, 2, ..., m\}$, *the* **flips within** K **applied to** s, $\phi_K(s) = \phi_{i_k}(...\phi_{i_2}(\phi_{i_1}(s))...)$.

Note that $\phi_K(s)$ doesn't depend on the order in which the flipsets are applied. The *span* of a set of flipsets is the set of all strings which can be generated from the null string s^0 by applying flipsets individually and in combination.

Def'n: *The* **span** *of* $A = \{\phi_1, ..., \phi_m\}$, $A^{\oplus} = \{s : s = \phi_K(s^0)$ *for some* $K \subseteq \{1, 2, ..., m\}\}$

The proposition that follows uses the elementary heuristic as a benchmark to view equivalence.

Proposition 6 *A perspective/heuristic pair* (M, A) *is equivalent to* (M', A_E^K) *for some perspective* M' *and some* $K \subseteq \{1, 2, ..., n\}$ *iff* $A^{\oplus}$ *has* 2^m *numbers of elements.*

Proposition 6 implies that when different perspectives are allowed, many heuristics (the ones that fully span m-dimensions) that are not elementary are equivalent to the elementary heuristics combined with appropriate perspectives. This raises the issue of the location of knowledge. Imagine two agents who perform the same searches on the space of objects. Agent 1 uses an elaborate perspective and the elementary heuristic, and Agent 2 uses a standard perspective and a sophisticated heuristic. In the former case, knowledge is embedded in the perspective. In the latter, the heuristic contains most of the agent's knowledge. Thus, to say that a perspective heuristic pair can be equivalent to an alternative perspective with a simple heuristic is *not* to say that any intelligent person can be replaced by a simpleton, but that some very sophisticated techniques can be recast as simple using different encodings. For example, graphing the function $sqrt 1 - x^2$ for x ranging from minus four to four is not trivial, but perfomring the same task in polar coordinates is trivial.

The next proposition states that there exist perspectives together with the elementary heuristic that are not equivalent to any heuristic paired with the identity perspective. This implies that there exist agents with perspectives and simple heuristics that cannot be imitated by agents using the identity perspective and sophisticated heuristics.

Proposition 7 *There exist a perspective* M *and a* $K \subseteq \{1, 2, ..., n\}$ *with* $|K| = m$ *s.t.* (M, A_E^K) *is not equivalent to* (I, A) *for any heuristic* A, *where* I *is the identity perspective.*

Throughout the paper, we have argued that modeling problem solving by differentiating between perspectives and heuristics has a great intuitive and realistic appeal. Here with Propositions 6 and 7, we have established that the specific way in which we model perspectives and heuristics as two dimensions of problem solving seems reasonable: that flexibility in heuristics can not replace flexibility in perspectives and vice-versa. This insight hinges on the assumption that heuristics here are string independent. Further, the result of Proposition 6 suggests that we can find a lower bound on the number of equivalence classes of all problem solvers. We do that next.

Proposition 8 *The number of equivalence classes is at least*

$$\sum_{m=1}^{n} \frac{2^n!}{2^{n-m}!(2^m m)^{2^{n-m}}}$$

For large n, this lower bound becomes astronomically large, demonstrating the enormous number of ways for people to encode and attempt to solve problems even when restricted to binary strings.

7 Discussion

In this paper we have constructed a model of economic agents with diverse problem solving behavior. The model distinguishes between perspectives, how people encode problems, and heuristics, how they go about trying to solve them. Our model provides underpinnings for analyses such as Grossman and Maggi (1999) that demonstrate the role of diversity in human capital in explaining growth and trade. Some of our findings, for example the notion that diversity leads to optimality, should delight economists. The fact that collections of agents of limited ability can solve difficult problems provides a stronger foundation for the optimality assumption. Other findings are less optimistic. We show that the path towards optimality exhibits little regularity. We demonstrate the possibility of arbitrary returns to adding workers of equal ability, and the possibility of both increasing and decreasing returns to additional workers with the same set of equal-ability workers.

The results in this paper rely on a binary string model **where agents flip bits. Some readers might take exception to this. Note thought that we are modeling agents confronted with problems that they cannot solve, such as curing cancer, developing workable forms of fusion, and disposing of radioactive wastes. Thus, it makes sense to make the agents appear simple relative to the problem at hand. That said, the question that should be asked of our model is whether the results would only occur in a binary string model or whether they can be generalized.** An examination of the proofs supports the generalizability of the claims. In order to generate the possibility of any monotonically increasing function as the total returns to problem solvers, the local optima of the various problem solvers must have basins of attraction which overlap in a particular way. This can be accomplished with any type of encoding. The same is true of the finding that order matters. Thus, the results seem quite general. Binary strings happen to provide the most convenient language for presenting the idea.

The interpretations so far have been limited to human problem solvers. Yet, the perspective/heuristic framework can equally well be applied to the special case of artificial problem solvers, such as computers. The mapping between our heuristics and computer algorithms is obvious. Further, computer searches rely on perspectives. To apply an algorithm to a problem, the problem must be encoded into a language that is natural for the computer. Two distinct computer algorithms may use identical or unique encoding. In carrying out this mental exercise of modeling computers within our framework, we uncover a difference between humans and computers as problem solvers. Humans possess common languages and enormous powers of visual interpretation which simplify communication. If people differ in their perspectives, and there would appear to be little basis for supposing that we all encode information identically, then the ability to encode visual stimuli offers an opportunity for the exploitation of this diversity. For computers, the use of diverse perspectives can be problematic. When a new best object is located by one computer, in order for the other computer to be informed of the object's identity, a look-up table must be consulted. This look-up table translates the object from one encoding into the other. Each look-up demands nontrivial computer time unless the encoding are related by some simple formula.

The current model is by no means complete. We offer it with the intention of spurring future research. Several extensions are apparent. In its present form the model does not allow for differences in the problem solvers' preferences over outcomes. These differences may stem from different incentives: one problem solver may not like a particular solution because he must work hard in that situation. Alternatively, different values may result from different beliefs about how policies map into outcomes. In politics, this often occurs. Both candidates offer growth, less crime, and a clean environment, but propose to achieve these ends through distinct policies. In either case, differences of opinion introduce a cost to increasing group size. Similarly, were we to include the possibility of miscommunication, this might bound the effective size of a group. Miscommunication might also offer an occasional improvement by dislodging search from local optima. Such modifications in the present model may enable us to generate insights about a variety of interesting questions including optimal group size as a function of problem difficulty.

References

[1] Arifovic, J. (1994). "Genetic algorithm learning and the cobweb model", *Journal of Economic Dynamics and Control* 18, 3-28.

[2] Arrow, K. and G. Debreu (1954). "Existence of equilibrium for a competitive economy", *Econometrica* 22, 265-290.

[3] Chatman, J., J. Polzer, S. Barsade, and M. Neale (1997). "Being Different Yet Feeling Similar: The Influence of Demographic Composition and Organizational Culture on Work Processes and Outcomes", mimeo, University of Texas at Austin.

[4] Cho, I. (1993). "Perceptrons Play Repeated Prisoner's Dilemma", mimeo, University of Chicago.

[5] Cyert, R. and J. March (1963). *A Behavioral Theory of the Firm,* Prentice Hall, Englewood Cliffs, New Jersey.

[6] de la Maza, M. and D. Yuret (1994). "A Futures Market Simulation with Nonrational Participants", in *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems.* R. A. Brooks and P. Maes eds, MIT Press.

[7] Debreu, G. (1959). *A Theory of Value,* Wiley, New York.

[8] Debreu, G. (1974). "Excess demand functions", *Journal of Mathematical Economics* 1, 15-21.

[9] Douglas, P. (1967). "Comments on the Cobb-Douglas Production Function", in *The Theory and Empirical Analysis of Production*, M. Brown ed., Columbia University Press.

[10] Frank, R. and P. Cook (1995). *The Winner-Take-All Society,* Free Press.

[11] Grossman, G. and G. Maggi (1999). "Diversity and Trade" mimeo, Department of Economics, Princeton.

[12] Hong, L. and S. Page (1998). "Diversity and Optimality", mimeo.

[13] Kalai, E. and E. Lehrer (1993). "Subjective Equilibrium in Repeated Games", *Econometrica* 61, 1019-1045.

[14] Kalai, E. and E. Lehrer (1995). "Subjective Games and Equilibria", *Games and Economic Behavior* 8, 123-163.

[15] Kalai, E. and W. Stanford (1988). "Finite rationality and interpersonal complexity in repeated games", *Econometrica* 56, 397-410.

[16] Kollman, K., J. Miller and S. Page (1992). "Adaptive Parties and Spatial Elections," *American Political Science Review* 86, 929–37.

[17] Knuth, D. (1968). *The Art of Computer Programming,* Addison–Wesley, Menlo Park, California.

[18] Macken, C., P. Hagan and A. Perelson (1990). "Evolutionary walks on rugged landscapes", Working Paper 90-007, Santa Fe Institute.

[19] Mantel, R., (1974). "On the characterization of aggregate excess demand", *Journal of Economic Theory* 7, 438-53.

[20] Marimon, R., E. McGrattan and T. Sargent (1990). "Money as a medium of exchange in an economy with artificially intelligent agents", *Journal of Economic Dynamics and Control* 14, 329-374.

[21] March, J. and H. Simon (1958). *Organizations,* John Wiley & Sons, New York.

[22] Marr, D. (1982). *Vision,* W. H. Freeman, New York.

[23] McKenzie, L. (1959). "On the existence of general equilibrium for a competitive market", *Econometrica* 27, 54-71.

[24] Milgrom, P. and J. Roberts (1992). *Economics, Organization and Management,* Prentice Hall, Englewood Cliffs, New Jersey.

[25] Miller, J. (1992). Ph. D dissertation, The University of Michigan–Ann Arbor.

[26] Newell, A. and H. Simon (1972). *Human Problem Solving,* Prentice Hall, Englewood Cliffs, New Jersey.

[27] Page, S. (1994). "Covers: A Theory of Boolean Function Decomposition," *Complex Systems,* 1-24.

[28] Page, S. (1996). "Two Measures of Difficulty," *Economic Theory*, 321-346.

[29] Penrose, R. (1989). *The Emperor's New Mind,* Oxford University Press, Oxford.

[30] Radner, R. (1992). "Hierarchy: The Economics of Managing", *Journal of Economic Literature* XXX, 1382-1415.

[31] Radner, R. (1993). "The organization of decentralized information processing", *Econometrica* 61, 1109-1146.

[32] Radner, R. and T. Van Zandt (1995). "Information Processing in Firms and Returns to Scale", in *The Economics of Informational Decentralization: Complexity, Efficiency, and Stability,* ed. John Ledyard, Kluwer Academic Publishers.

[33] Robbins, S. (1994). *Organizational Behavior,* Prentice Hall, Saddle River, New Jersey.

[34] Rubinstein A. (1986). "Finite automata play the repeated prisoner's dilemma", *Journal of Economic Theory* 39, 83-96.

[35] Sargent, T. (1993). *Bounded Rationality in Macroeconomics*, Oxford.

[36] Sonnenschein, H. (1973). "Do Walras' identity and continuity characterize the class of community excess demand functions?", *Journal of Economic Theory* 6, 345-54.

[37] Thomas, D. and R. Ely (1996). "Making Differences Matter: A New Paradigm for Managing Diversity", *Harvard Business Review* #96510.

[38] Watson, W., K. Kumar, and L. Michaelsen (1993). "Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups", *Academy of Management Journal* 36, 590-602.

Appendix 1

In this appendix, we describe the sources of our data on the job classification of the workforce in years 1950, 1970 and 1992 and list in detail jobs that we categorize under the heading problem solvers.

The data for 1950 are from *The Statistical History of the United States* published in 1976. The data for 1970 and 1992 are from *Statistical Abstract of the United States* published in 1975 and 1993 respectively. The job classification changes from year to year. The 1950 data we use are according to 1960 classification provided in the above mentioned source.

Following is the list of jobs we include in the category of problem solvers:

- *Problem Solvers*

  Executive, Administrative, and Managerial Specialty

  officials and Administrators, public
  financial managers
  personnel and labor relations managers
  purchasing managers
  managers, marketing, advertising and public relations
  administrators, education and related fields
  managers, medicine and health
  managers, properties and real estate
  management related occupations such as accountants and auditors

  Professional Specialty

  architects
  engineers
  mathematical and computer scientists
  computer Programmers
  natural scientists such as chemists and biological scientists
  health diagnosing occupations such as physicians and dentists
  social scientists and urban planners
  lawyers and judges
  designers
  public relations specialists

Appendix 2

We begin by proving Propositions 3 and 4.

Proposition 3 *For any $m \geq 3$, there exist an $n \geq 1$, a value function V defined on binary strings of length n, a group of problem solvers $\{P_i\}_{i=1}^m$, and σ, σ', two elements of the permutation group on $\{1, ..., m\}$, such that (i),(ii), and (iii) hold:*

(*i*) $E[V : \{P_{\sigma(j)}\}_{j=1}^{i}] - E[V : \{P_{\sigma(j)}\}_{j=1}^{i-1}] \geq E[V : \{P_{\sigma(j)}\}_{j=1}^{i-1}] - E[V : \{P_{\sigma(j)}\}_{j=1}^{i-2}]$ for $i = 3, ..., m$

(*ii*) $E[V : \{P_{\sigma'(j)}\}_{j=1}^{i}] - E[V : \{P_{\sigma'(j)}\}_{j=1}^{i-1}] \leq E[V : \{P_{\sigma'(j)}\}_{j=1}^{i-1}] - E[V : \{P_{\sigma'(j)}\}_{j=1}^{i-2}]$ for $i = 3, ..., m$

(*iii*) $E[V : P_i] = E[V : P_j]$ for all $i, j = 1, ..., m$

Pf: In the proof we rely on the case $m = 5$. The proof for a general m follows the $m = 5$ case exactly. Choose $n = 1+2+...+(m-1) = \frac{m\cdot(m-1)}{2}$.. When $m = 5$, $n = 10$ – the sum of the integers 1 through 4. The domain of the value function consists of 2^n object strings. Let one object string have a value of 1, $2n$ objects have a value of -1 and all other object strings have a value of -2. Assume that every problem solver uses the elementary heuristic. The perspectives for the agents are as follows. Let x^1 through x^{2n} denote the $2n$ object strings with value -1 and x^0 denote the object string of value 1.

Let s^0 denote the string of all zeroes. We assume that all agents encode x^0 as s^0. Let s^{1i} for $i = 1$ to n denote the n strings with exactly one bit set equal to 0, and let s^{0j} for $j = 1$ to n denote the n strings with exactly one bit set equal to 1. Notice that the s^{0j}'s lie in the basin of attraction of the optimal string s^0, while the s^{1i}'s are local optima with respect to the elementary heuristic.

In each of the m problem solver's perspective, these $2n$ strings denote the $2n$ object strings of value -1. The perspectives only differ in their mappings from the x^i's to the s^{1i}'s and the s^{0j}'s. Define the perspective of the mth problem solver as follows:

$$M_m(x^i) = \begin{cases} s^{1i} & \text{for } i = 1 \text{to } n \\ s^{0(i-n)} & \text{for } i = n+1 \text{to } 2n \end{cases}$$

In the case of $m = 5$, the globally optimal object, x^0 and the ten objects x^1 to x^{10} are local optima for the fifth problem solver. In what follows, each problem solver has the same local optima which have value -2. They affect the expected value in exactly the same way for each group of problem solvers. For ease of exposition, we don't keep track of the local optima individually.

We need one more piece of notation. For $k = 1$ to $m-1$, let Σ_k be the sum of the integers from 1 through k. Define the perspective of problem solver 1 as follows:

$$M_1(x^i) = \begin{cases} s^{0i} & \text{for } i = 1 \text{to } \Sigma_1 \\ s^{1(i-\Sigma_1)} & \text{for } i = \Sigma_1 + 1 \text{to } \Sigma_1 + n \\ s^{0(i-n)} & \text{for } i = \Sigma_1 + n + 1 \text{to } 2n \end{cases}$$

Therefore, in the case $m = 5$, the local optima for problem solver 1 are $\{x^0, x^2, x^3, x^4, x^5, x^6, x^7, x^8, x^9, x^{10}, x^{11}\}$. The set of local optima for the group consisting of problem solver 1 and problem solver 5 consists of x^0 and x^2 through x^{10}. The perspectives of problem solvers k for $k = 2$ to $m - 1$ are defined as follows:

$$M_k(x^i) = \begin{cases} s^{1i} & \text{for } i = 1 \text{to } \Sigma_{k-1} \\ s^{0(i-\Sigma_{k-1})} & \text{for } i = \Sigma_{k-1} + 1 \text{to } \Sigma_k \\ s^{1(i-k)} & \text{for } i = \Sigma_k + 1 \text{to } k + n \\ s^{0(i-n)} & \text{for } i = k + n + 1 \text{to } 2n \end{cases}$$

In the $m = 5$ case, the local optima for problem solvers 2 and 3 respectively are $\{x^0, x^1, x^4, x^5, x^6, x^7, x^8, x^9, x^{10}, x^{11}, x^{12}\}$ and $\{x^0, x^1, x^2, x^3, x^7, x^8, x^9, x^{10}, x^{11}, x^{12}, x^{13}\}$. These sets are important in the proof. When problem solver 2 is added to the group consisting of problem solvers 1 and 5, the set of local optima for the group is reduced by two. And when problem solver 3 is added to the group consisting of problem solvers 1,2, and 5, the set of local optima for the group is reduced by three. Note that the reduction of each such local optimum increases the expected value by the same amount. Denote this amount by δ. By construction, $\delta > 0$. If the order in which problem solvers are added to the group is 5,1,2,3,4, then the expected value of adding problem solver i equals $i \cdot \delta$. So, we have increasing returns to adding problem solvers. But by construction, if the order in which the problem solvers are added to the group is 5,4,3,2,1, then the expected value of adding problem solver i is still $i \cdot \delta$ and we have decreasing returns to adding problem solvers. All that remains to see is that the expected value of local optima for each problem solver is the same. But this follows from the construction.

We next prove Proposition 4 which shows that any weakly increasing function can be generated as the performance of diverse problem solvers of equal ability.

Proposition 4 *For any $m \geq 2$ and any function $f : \{1, 2, ..., m\} \to Q$ (Q denotes the set of rational numbers) which is weakly increasing with $f(1) = 0$ and $f(m) = 1$, there exist an $n \geq 1$, a value function $V : \{0,1\}^n \to \Re$, and problem solvers $\{P_i\}_{i=1}^m$ such that $E[V : \{P_i\}_{i=1}^j] = f(j)$ for $j = 1$ to m and that $E[V : P_i] = E[V : P_j]$ for all i and j.*

Pf: We assume a strictly increasing function. The weakly increasing case follows immediately by adding problem solvers identical to earlier problem solvers. Let $\aleph$ be the set of positive integers. For $i = 2, ..., m$, let $d(i) = f(i) - f(i-1)$. We define n as follows:

$$n = 3 \cdot min\{x : x \in \aleph,\ \exists k_i \in \aleph \text{ s.t. } \frac{k_i}{x} = d(i) \text{ for all } i = 2, ..., m\}$$

We construct the value function similar to the one used in the proof of Proposition 3. As before, the domain of the value function consists of 2^n object strings, and s^0 denotes the string of all 0's. Further, s^1 denotes the string of all 1's, s^{1i}'s for $i = 1, ..., n$ denote n strings with exactly one bit equal to 0, and s^{0j}'s for $j = 1, ..., n$ denote n strings with exactly one bit equal to 1. For each object string s, we define the value function $V : S \to \Re$ as follows:

$$V(s) = \begin{cases} 1 & \text{for } s = s^0 \\ -2 & \text{for } s = s^1 \\ -\min\ (\Sigma_{j=1}^n s_j,\ n - \Sigma_{j=1}^n s_j) & \text{for} s \neq s^0, s^1 \end{cases}$$

As before, we denote object strings s^{1i}'s for $i = 1, ..., n$ by x^1 through x^n and object strings s^{0j}'s for $j = 1, ..., n$ by x^{n+1} through x^{2n}. By the definition of the value function V above, $V(x^i) = -1$ for $i = 1$ to $2n$. The m problem solvers use the elementary heuristic. Their perspectives differ only in the way they encode object strings x^1 through x^{2n}. They are defined below. For $k = 2, ..., m$, let $\Sigma_k = nf(k)$.

$$M_1(x^i) = \begin{cases} s^{1i} & \text{for } i = 1 \text{to } n \\ s^{0(i-n)} & \text{for } i = n+1 \text{to } 2n \end{cases}$$

$$M_2(x^i) = \begin{cases} s^{0i} & \text{for } i = 1 \text{ to } \Sigma_2 \\ s^{1(i-\Sigma_2)} & \text{for } i = \Sigma_2 + 1 \text{ to } \Sigma_2 + n \\ s^{0(i-n)} & \text{for } i = \Sigma_2 + n + 1 \text{to } 2n \end{cases}$$

For $k = 3, ..., m$,

$$M_k(x^i) = \begin{cases} s^{1i} & \text{for } i = 1 \text{to } \Sigma_{k-1} \\ s^{0(i-\Sigma_{k-1})} & \text{for } i = \Sigma_{k-1} + 1 \text{to } \Sigma_k \\ s^{1(i-\Sigma_k+\Sigma_{k-1})} & \text{for} i = \Sigma_k + 1 \text{to } \Sigma_k - \Sigma_{k-1} + n \\ s^{0(i-n)} & \text{for } i = \Sigma_k - \Sigma_{k-1} + n + 1 \text{ to } 2n \end{cases}$$

For each problem solver $k = 1, ..., m$, the set of local optima consists of s^0 and s^{1i}'s for $i = 1, ..., n$. These are M_k-strings. By a simple calculation, the probability of the search ending at a given s^{1i} equals $\frac{1}{2n}$, and the probability of the search ending at s^0 is $\frac{1}{2}$.. Therefore, we have $E[V : P_i] = E[V : P_j] = 0$ for all i and j. In particular, $E[V : P_1] = 0 = f(1)$. A straightforward calculation shows that the expected value of a local optima for problem solvers 1 and 2 equals $0 + \frac{\Sigma_2}{n} = f(2)$. Similarly, for $j = 3$ to m, $E[V : \{P_i\}_{i=1}^j] = f(j-1) + \frac{\Sigma_j - \Sigma_{j-1}}{2n} \cdot 2 = f(j)$.

We next prove the three Propositions in Section 6 regarding equivalence. A more detailed characterization of combining flipsets is required. If $A = \{\phi_1, ..., \phi_m\}$ and

$K = \{1, 3, 4\}$, then ϕ_K represents a flipset created by applying ϕ_1, followed by ϕ_3, and finally ϕ_4. For example, if $\phi_i = \{i, i+1\}$, then applying $\phi_{\{1,3,4\}}$ to a string amounts to flipping the first and the second bits and then flipping the third and the fourth bits followed by flipping fourth and fifth bits. The strings are defined over a binary alphabet, so flipping the fourth bit twice is equivalent to not flipping it at all. Thus, $\phi_{\{1,3,4\}} = \{1, 2, 3, 5\}$. To take into account the fact that flipping a bit an even number of times is the same as not flipping the bit, we define the operator $\oplus$.

Def'n: *The operator* $\oplus : \wp(\{1, 2, ..., m\}) \times \wp(\{1, 2, ..., m\}) \rightarrow \wp(\{1, 2, ..., m\})$ *according to the following rule:* $J \oplus K = (J \cup K) \setminus (J \cap K)$ *for* $J, K \subseteq \{1, 2, ..., m\}$ *where* $\wp(\{1, 2, ..., m\})$ *is the power set of* $\{1, 2, ..., m\}$.

Example: $\{1,\ 3,\ 4\} \oplus \{4,\ 5\} = \{1,\ 3,\ 5\}$.

Remark 1 *Given any heuristic* $A = \{\phi_1, ..., \phi_m\}$ *and* $K, J \subseteq \{1, 2, ..., m\}$, $\phi_J(\phi_K(s)) = \phi_{J \oplus K}(s)$ *for all* $s \in S$.

Proposition 6 *A perspective/heuristic pair* (M, A) *is equivalent to* (M', A_E^K) *for some perspective* M' *and some* $K \subseteq \{1, 2, ..., n\}$ *iff* $A^{\oplus}$ *has* 2^m *numbers of elements.*

To prove Proposition 6, we only need to show the following two lemmas.

Lemma 1 (I, A) *is equivalent to* (M_1, A_E^K) *for some perspective* M_1 *and some* $K \subseteq \{1, 2, ..., n\}$ *iff* $A^{\oplus}$ *has* 2^m *numbers of elements.*

Pf: $(\Rightarrow)$ Since $|A| = m$, $|K| = m$. It is not difficult to show (we omit it) that $M_1(A^{\oplus}) = A_E^{K\oplus}(M_1(s^0))$ where $A_E^{K\oplus}(M_1(s^0))$ denotes the set of strings that can be reached from $M_1(s^0)$ by applying flipsets in A_E^K individually and in combination. Obviously $A_E^{K\oplus}(M_1(s^0))$ has 2^m elements. Since M_1 is 1-to-1, $A^{\oplus}$ has 2^m elements.

$(\Leftarrow)$ When $A^{\oplus}$ has 2^m elements, it can be shown that there exists a set of 2^{n-m} number of strings including s^0, $\{s^0, s^1, ..., s^{2^{n-m}-1}\}$, s.t. (a) $S = \cup_{j=0}^{2^{n-m}-1} A^{\oplus}(s^j)$ and $A^{\oplus}(s^j) \cap A^{\oplus}(s^k)$ for any $j \neq k$ where $A^{\oplus}(s^j)$ denotes $\{s \in S : s = \phi_J(s^j)$ for some $J \subseteq \{1, ..., m\}\}$ (notice $A^{\oplus}(s^0) = A^{\oplus}$) (b) for any $s \in A^{\oplus}(s^j)$, there exists a unique subset of $\{1, ..., m\}$, denote it $J(s)$, s.t. $s = \phi_{J(s)}(s^j)$. Now we define M_1 as follows: for any $s \in A^{\oplus}(s^j)$, $M_1(s)$ is the unique string such that $\{i \in \{1, ..., m\}$: the ith bit of $M_1(s)$ differs from the ith bit of $s^j\} = J(s)$. Let $K = \{1, ..., m\}$. We can then easily show that for any given $s \in S$, $M_1(\phi_k(s)) = \phi_k^e(M_1(s))$ for any $k \in \{1, ..., m\}$. Thus (I, A) is equivalent to (M_1, A_E^K).

Lemma 2 *(a) If* (M, A) *is equivalent to* (M', A_E^K) *for some* $K \subseteq \{1, 2, ..., n\}$, *then* (I, A) *is equivalent to* (M_1, A_E^K) *where* $M_1 = M' \circ M^{-1}$. *(b) If* (I, A) *is equivalent to* (M_1, A_E^K) *for some* $K \subseteq \{1, 2, ..., n\}$, *then for any perspective* M, (M, A) *is equivalent to* (M', A_E^K) *where* $M' = M_1 \circ M$.

Pf: We only prove (a) here, (b) can be similarly proven. Obviously, $|K| = m$. To show that (I, A) is equivalent to (M_1, A_E^K), we need to show for any $s \in S$, there exists a 1-1 map $\sigma : \{1, 2, ..., m\} \rightarrow K$ such that $\phi_k(s) = M_1^{-1}(\phi^e_{\sigma(k)}(M_1(s)))$. Fix any $s \in S$, consider $s' = M^{-1}(s)$. Since (M, A) is equivalent to (M', A_E^K), we know that for s' there exists a 1-1 map $\sigma : \{1, 2, ..., m\} \rightarrow K$ such that $M^{-1}(\phi_k(M(s'))) = M'^{-1}(\phi^e_{\sigma(k)}(M'(s)))$. This implies $\phi_k(s) = M_1^{-1}(\phi^e_{\sigma(k)}(M_1(s)))$ since $s' = M^{-1}(s)$ and $M_1 = M' \circ M^{-1}$.

Proposition 7 is much simpler. It says

Proposition 7 *There exist a perspective M and a $K \subseteq \{1, 2, ..., n\}$ with $|K| = m$ s.t. (M, A_E^K) is not equivalent to (I, A) for any heuristic A.*

Pf: We construct an example for the case $n = m = 3$, which can be extended to an arbitrary n. Consider the perspective M.

s	0	1	2	3	4	5	6	7
$M(s)$	0	1	2	7	4	5	6	3

We claim that there does not exist a heuristic A s.t. (I, A) is equivalent to (M, A_E). The object string 0 and the object strings 1, 2, and 4 are mapped into themselves under the perspective M. Strings 1,2, and 4 are the neighbors of string 0 using the elementary heuristic. Therefore, if (I, A) is equivalent to (M, A_E), the neighbors of object string 0 must be object strings 1,2, and 4, and A must be the elementary heuristic. But I and the elementary heuristic together is not equivalent to (M, A_E).

For each problem to be solved, we can look at the number of equivalence classes among all the problem solvers. Proposition 6 helps us to establish a lower bound on the number of equivalence classes.

Def'n: $\Pi = \{(M, A) : M$ *is a perspective and* $A = \{\phi_1, ..., \phi_m\}$ *is a flipset heuristic such that* $A^{\oplus}$ *is isomorphic to an* m*-dimensional hyperplane of* S *for some* m*,* $1 \leq m \leq n\}$

Proposition 8 *The number of equivalence classes is at least*

$$\sum_{m=1}^{n} \frac{2^n!}{2^{n-m}!(2^m m)^{2^{n-m}}}$$

Pf: We only need to show that there are $\sum_{m=1}^{n} \frac{2^n!}{2^{n-m}!(2^m m)^{2^{n-m}}}$ many equivalence classes in Π. By Proposition 6, we know that for any $P = (M, A) \in \Pi$ where $A = \{\phi_1, ..., \phi_m\}$, there exists a perspective $M' : S \to S$ and $K \subseteq N$ with $|K| = m$ such that (M, A) is equivalent to (M', A_E^K) which also belongs to Π. Thus we only need to consider the number of equivalence classes in $\{(M, A_E^K) : M$ is a perspective and $K \subseteq N\}$. Denote $\{(M, A_E^K) : M$ is a perspective and $K \subseteq N\}$ by Π_E. Obviously, if $|K| \neq |K'|$, then for any two perspectives M and M', (M, A_E^K) and $(M', A_E^{K'})$ are not equivalent. Therefore, the number of equivalence classes in $\Pi_E = \sum_{m=1}^{n}$the number of equivalence classes in Π_E^m, where $\Pi_E^m = \{(M, A_E^K) : M$ is a perspective and $K \subseteq N$ with $|K| = m\}$. The following lemma proven at the end leads to the result.

Lemma 3 *Given a perspective M, and K, $K' \subseteq N$ such that $|K| = |K'|$, there exists another perspective M' such that (M, A_E^K) is equivalent $(M', A_E^{K'})$.*

From Lemma 3, we know that the number of equivalence classes in Π_E^m is equal to the number of equivalent classes in Π_E^K where $K \subseteq N$ and $|K| = m$ is arbitrarily fixed and $\Pi_E^K = \{(M, A_E^K) : M$ is a perspective$\}$. It can be shown (we leave out the details) that for any fixed M, there are

$$(2^m m)^{2^{n-m}} (2^{n-m}!)$$

many (M', A_E^K)'s that are equivalent to (M, A_E^K). Since there are total of $2^n!$ many elements in Π_E^K, the number of equivalence classes in Π_E^K is

$$\frac{2^n!}{2^{n-m}!(2^m m)^{2^{n-m}}}$$

Therefore, the number of equivalence classes in Π is

$$\sum_{m=1}^{n} \frac{2^n!}{2^{n-m}!(2^m m)^{2^{n-m}}}$$

Pf of Lemma 3: By Lemma 2, we only need to show that (I, A_E^K) is equivalent $(M', A_E^{K'})$ for some perspective M'. Let $\sigma : N \to N$ be a 1-1 map such that $\forall i \notin (K \cup K') \backslash (K \cap K')$, $\sigma(i) = i$. Define a perspective M' as follows: for any $s \in S$, $M'(s) = s_{\sigma(n)} \cdots s_{\sigma(1)}$. Then, it is straightforward to show that (I, A_E^K) is equivalent $(M', A_E^{K'})$.