Problem Solving by Heterogeneous Agents

Lu Hong[*] and Scott E. Page[†‡]

October 22, 1998

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 confronting difficult problems 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.* With this model, we find that a collection of bounded but diverse agents can locate optimal solutions to very difficult problems. We can also calculate the marginal benefits to adding additional problem solvers. We find that problem solving firms can exhibit arbitrary returns to scale, that the order that problem solvers are applied to a problem can matter, and that the standard story of decreasing returns to scale is unlikely.

1 Introduction

Economics has, by and large, treated workers as having unidimensional ability. In effect, this reduces humans to robots of various speeds. This assumption requires either that people do not differ in our behavior, some of us are just smarter or dumber than others, or that economically relevant activities require workers to perform standard and repetitive tasks, that economic behavior is robotic. Forty years ago, when a majority of workers produced goods and services (when they made stuff), this assumption made sense. The models when tested proved accurate. The Cobb–Douglas

[*]Department of Economics, Syracuse University, Syracuse, NY 13244. lhong@mailbox.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, Midwest Math. Economics Meeting in Minneapolis, the workshop on New Directions in the Theory of Markets and Games, 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.

production function was originally an empirical result.[1] Identical workers performing repeated, standard tasks exhibit diminishing marginal products. The implications for markets and general equilibrium theory 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. More people provide services, process information (Radner 1993), solve problems, and work for the government than manufacture goods. And the manufacturing sector shrinks daily. It stands to reason that the canonical general equilibrium model of economic activity might benefit from amending, or at a minimum from a reinterpretation of the foundations. In many cases, the neoclassic model of Arrow-Debreu (1954), Debreu (1959) and McKenzie (1959) remains an accurate representation, particularly for the provision of services. Information processing, however, appears to differ slightly (Radner and Van Zandt 1995), although decreasing marginal returns can be established given reasonable assumptions.

To our knowledge, no models exist that analyze the remaining type of firms. Firms that solve problems. Firms that search for cures to diseases, that develop software, that design homes and bridges, that handle legal cases, that produce research, that develop pollution reducing technology, that design welfare policies, that make movies, that design video games, that engineer new drugs. 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 activities including manufacturing (where workers perform standard tasks), based on which decreasing marginal product of labor is established and many results of the standard economic model, such as wage equals marginal product of labor, are drawn. How applicable are these results to problem solving? To address this question, we need a model of problem solving that takes into account human ingenuity. The model must capture the essence of problem solving and reveal how it differs from other production activities. Second, the model must account for the fact that despite conflicting interests and/or miscommunication, groups and not individuals most often attempt difficult problems. At its core, any such model must include assumptions that explain why groups outperform individuals. This paper represents our attempt to construct such a model.

Any model of problem solving agents must abandon the perfect rationality assumption, lest problem solving becomes trivial, not to mention that the assumption is unrealistic. No one knows how to cure the common cold, let alone fold proteins. The problem solvers must have limited abilities. They must also be allowed to differ in the ways they encode and approach problems. This approach not only allows the capture of individuality but also provides the basis for an explanation of why collective effort by a group can often outperform an individual: by virtue of being different, individuals can improve upon each other's solutions to a problem.

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

In the formal model, we assume that every solution to the problem can be evaluated and 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. We assume also that the problem is difficult. Agents are unable to simply choose the optimal solution given their limited ability. Instead, they find the best solutions they can. In the 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 can identify 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 in economics, it is not without foundation. In their book *Human Problem Solving*, Newell and Simon (1972) studied human problem solving and explicitly modelled the process as following two steps: representation of the problem and then applying a heuristic, although not necessarily only once in the process. Moreover in recent years, many artificial agent models of decision making have been introduced into economics. They can be found in macroeconomics (Sargent 1993), game theory (Kalai and Lehrer 1993 & 1995), and political economy (Kollman, Miller, and Page 1993). 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 perspective which is chosen by the modeler and is fixed. Our model can be seen as a generalization that contains many of these specific examples.

As mentioned above, one distinction between our model and these artificial agent models is that we allow agents to have different perspectives. Adding this dimension provides a potentially richer description of human capital and may enlarge the scope of diversity in problem solving. Such diversity is our explanation for the prevalence of group problem solving as opposed to individual problem solving. By including the dimension of perspectives, we encompass as much as possible diverse problem solving and at the same time keep in line with the cognitive foundation of human problem solving. The logic of the argument here relies on the crucial assumption that introducing perspectives *truly* enlarges the set of all different ways of solving a problem compared to if one fixes a perspective and only allows heuristics to vary. However, this assumption is by no means trivial and needs to be justified. In the paper, we explore in great detail a special problem solving model, the binary string model, which has more structure than the general model, and justify this assumption by establishing a number of theorems that rely on the concept of equivalence classes. Two agents belong to the same equivalence class if their distinct 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 we find for this more structured environment becomes astronomical as the problem size gets larger. 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 the carefully constructed binary string model of collective problem solving, we derive two categories of results. The first category of results 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 a difficult problem. 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. Nothing in our analysis requires that the agents make a decision as a group. All of the agents may work in isolation. It is only that they have a collective goal.

The second category of results relies on interpreting our model as representative of problem solving within firms. We calculate the returns to additional workers, the improvement in the value of the best solution brought about by adding a worker and examine their properties. Here, we arrive at some rather interesting conclusions. We prove the possibility of arbitrary returns to additional problem solvers, and 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. However, as we shall argue, these conditions defy basic economic logic. The agents must sort into problems that confound them. Taken together, our results suggests 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 missing from our model. We ignore asymmetric and imperfect information. Every agent can compute the value of a solution. They all use the same value function. We ignore incentives. This assumption may be problematic given the importance placed on incentive constraints in organizational structure and performance (Milgrom and Roberts 1992). Relaxing either of these assumptions would allow for differences in values of objects opening the door to preference cycles, agenda manipulation and other difficulties. We ignore communication problems. Solutions can be costlessly and errorlessly communicated to other agents. Finally, we ignore the obvious larger implications of our modelling framework. The findings implicitly critique context free, unidimensional measures of ability. We discuss this delicate issue in a separate paper (Hong and Page 1998).

We have organized the remainder of this paper in seven parts. In section 2, we provide some data describing the number of problem solving workers in the U.S. economy and we comment on our assumption of diversity in human capital. 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.[2] In section 4, we describe in some detail a 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 5, we establish the aforementioned equivalence relationship among perspective/heuristic pairs. In section 6, we prove a fundamental link between diversity among bounded agents and collective optimality. In section 7, we examine 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. We conclude with a discussion of the robustness of our results as well as comments on some related issues including problem solvers who differ in their interpretations of the values of objects, the possibility of miscommunication between problem solvers, and differences between human 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 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 gather data from the United States to support the first assumption. The second assumption seems noncontroversial. Fresh perspectives and new ideas provide the basis for many improvements in problem solving contexts. We cite some recent studies and elaborate further on this intuition.

2.1 Problem Solvers: The Data

Calculating the percentage of workers who should be classified as problem solvers is a difficult task, so before entering into a micro–level discussion of job classifications, we begin by reporting some aggregate statistics which strongly support our first assumption. While it is true that 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 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 classify equally well as problem solvers.

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

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.[3] 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 in great detail 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 could be considered problem solving. For example, the numbers include management consultants, lawyers, and computer programmers but exclude bankers and teachers. See Appendix 1 for a complete description of those job categories included as problem solvers.

The Increase in Problem Solvers

Year	**1950**	**1970**	**1992**[4]
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

[5]

The data show a substantial increase in the percentage of problem solvers in the economy over the last forty years.

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

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

2.2 Improvement Through Diversity

The second assumption, that heterogeneity of perspectives and heuristics explains the benefits of collections of agents, requires a more subtle motivation. There are several reasons why groups may outperform individuals. The theory presented here encompasses many of them. Consider, for example, the specialization explanation: groups are better at solving multidimensional problems because individuals can develop expertise on components of the larger problem. Our model accounts for this phenomenon – the heuristics of the agents can be constructed so as to apply to a portion of the domain, thereby mimicking specialization. We do not wish to dedicate much effort to summarizing existing theories of group performance other than to say that our assumption supports many and does not appear to conflict with others.

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. 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 that 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 fact, Thomas and Ely (1996) go so far as to say 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 General Model

We begin by constructing a general model of a finite group of problem solvers of limited ability attempting to maximize a given value function defined over a large but finite set of objects X (potential solutions to the problem). All problem solvers agree on the value of the objects given by a value function $V : X \to \Re$. The problem solvers collectively attempt to find an optimal (or satisfactory) solution to the problem.

Each problem solver consists of an *internal language*, a *perspective*, and a *heuristics*. A problem solver uses her internal language to represent the objects. It is denoted by Γ. 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. More formally, a problem solver P is characterized by a dichotomy (M, A), where M denotes her perspective and A denotes her heuristic.

Def'n: *A* **perspective** M *of a problem solver* P *is a 1-1 mapping from the set of objects* X *to the problem solver's internal language* Γ*, i.e.,* $M : X \to \Gamma$ *is 1-1. Let* $S = M(X) \subset \Gamma$*. 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* $k = 1, 2, ..., m$*,* $\phi_k : S \to S$*.*

The internal language may be interpreted at either of two levels. Neurologically, our brains perceive and store information, and these perceptions differ across individuals. Metaphorically, people interpret problems based on their training, be it as economists, lawyers, etc.. A perspective may not be defined over all of X, so the problem solver need not be able to represent all objects in her internal language. A perspective also may be many to one. More than one object are mapped to the same representation in the internal language.[6]

The following example illustrates the concepts of perspective and heuristic.

Example 1: a job assignment problem Suppose n workers are to be assigned to m jobs. Assume $n \geq m \geq 2$. Different assignments lead to different levels of profit. A problem solver searches for an assignment of workers to jobs that generates the highest profit.

One natural representation of this problem is a set of $m \times n$ matrices. In particular, each assignment is represented by a matrix, $(a_{ij})_{i=1,...,m;j=1,...,n}$ where $a_{ij} = 0$ or 1, and $a_{ij} = 1$ if and only if worker j is assigned to job i. A heuristic on such a set of $m \times n$ matrices can be a collection of rules that specify switches of pairs of rows or pairs of columns.

A heuristic consists of rules for adapting the status quo solution in the problem solver's perception which might lead to improvement. It can be thought of as a

[6]See Kirschenheiter (1993) for an intriguing model which subsumes perspectives along these lines.

problem solver's bag of tricks. A problem solver searches from any status quo $s \in S$ 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 immediately to mind. First, $\phi_j(s)$ becomes a new status quo if and only if it gives the highest value[7] among all the $\phi_k(s)$'s and it has a strictly higher value than s. Second, apply the ϕ_k to the status quo object s, the first $\phi_k(s)$ that has a strictly higher value than s becomes a new status quo. The results presented in this paper do not depend on the exact search process a problem solver uses. We assume the use of this second type of sequential search process. In both cases, search stops at a solution s in the problem solver's internal language if and only if there is no further improvement by applying any of the ϕ_k's. The object in the original domain X that corresponds to such s is naturally called a *local maximum*. Specifically,

Def'n: *Given a problem solver $P = (M, A)$ where $A = \{\phi_1, \phi_2, ..., \phi_m\}$. The* **neighborhood of an object** $x \in X$ **for** P, *denoted by $b_M(A, x)$, consists of all the objects in X whose corresponding objects in P's internal language are the results of applying heuristic A to the x's corresponding object in P's internal language, i.e.,*

$$b_M(A, x) = \{x' : M(x') = \phi_k(M(x)) \text{ for some } k = 1, 2, ..., m\}$$

An object $x \in X$ is a **local maximum of** V **with respect to** P, *denoted by $x \in L(P, V)$, if and only if*

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

In modelling a group of problem solvers working to locate a solution to a common value function, we postulate that problem solvers have their own perspective/heuristic pairs. In particular, if they have expertise in locating a good solution, it is incorporated in their perspective/heuristic pairs. With this postulation, a team has 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. Of course, different perspective/heuristic pairs do not necessarily mean that they locate different solutions. In fact, one can imagine two problem solvers with distinct perspective/heuristic pairs creating identical neighborhoods and thus the same set of local maxima. 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. In later sections, we discuss in depth the issues of equivalence and diversity of problem solvers and how diversity contributes to better performance in a binary string model, a special case of our general model, where agents rely on the same internal language - binary strings - according to which they encode objects, and where heuristics are collections of flipsets. Here we 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 better solution than individual agents separately.

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

Example 2: A team of city council members try to locate a solution to the following public goods provision problem: 3 public projects, $\{p_1, p_2, p_3\}$, are under consideration (for example, build a subway route between downtown and the lake, clean up the lake, build a public library by the lake). The issue is which one(s), if any, of the projects to fund. There are 8 potential solutions. The net value for the city of each solution is described in the following table. The goal of the team is to locate a solution that generates the highest value for the city.

objects (possible solutions)	values
x_0: fund none	0
x_1: fund p_1 only	40
x_2: fund p_2 only	20
x_3: fund p_1 and p_2	60
x_4: fund p_3 only	30
x_5: fund p_1 and p_3	50
x_6: fund p_2 and p_3	70
x_7: fund all three projects	10

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 each *bit* s_i $(i = 1, 2, 3)$ can take a value of either 1 or 0; for her, 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 following table. Her heuristic A_1 consists of three functions denoted by $A_1 = \{\{1\}, \{2\}, \{3\}\}$. Each function $\{i\}$ $(i = 1, 2, 3)$ maps a binary string $s = s_3s_2s_1$ into a new string by flipping the ith bit s_i from 0 to 1 if $s_i = 0$ or 1 to 0 if $s_i = 1$. Intuitively, it means she knows three tricks, applying trick $\{1\}$ will mean to change the decision about p_1 only.

x	$M_1(x)$
x_0: fund none	000
x_1: fund p_1 only	001
x_2: fund p_2 only	010
x_3: fund p_1 and p_2	011
x_4: fund p_3 only	100
x_5: fund p_1 and p_3	101
x_6: fund p_2 and p_3	110
x_7: fund all three projects	111

Let's look at a possible search path of Member 1. She starts at x_0 (fund no projects) which she encodes into 000 and its value equals 0. She updates status quo string by applying her tricks sequentially, say in the order of $\{1\}, \{2\}$,and $\{3\}$.

Applying trick $\{1\}$ to the string 000 will lead her to 001 which corresponds to x_1 (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 trick $\{2\}$ to the string 001 to get 011 which corresponds to x_2 (fund p_1 and p_2) with value 60. The status quo is then updated to 011. Applying trick $\{3\}$ to 011 leads to 111 which corresponds to x_7 (fund all projects) and has a value 10. No updating occurs this time. She then applies trick $\{1\}$ to the string 011 to get 010 which corresponds to x_2 (fund p_2 only) and has a value 20 and thus no update. Applying trick $\{2\}$ to the string 011 to get 001 which corresponds to x_1 (fund p_1 only) and has a value 40 and thus again no update. Therefore, there is no improvement applying any of the three tricks to the string 011 and the search ends and the solution of such a search is x_3 (fund p_1 and p_2) with value 60. x_3 is a local maximum for Member 1. It is easily seen that the set of local maxima for Member 1 with (M_1, A_1) is $\{x_3, x_5, x_6\}$ in which x_6 is the maximum of the value function. Member 1 with her perspective/heuristic pair will always end up with one of the three local maxima. The initial object and the order that the tricks are applied will determine the local optimum that is obtained.

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. Her perspective M_2 is summarized in the following table. Her heuristic A_2 consists of only two functions denoted by $A_2 = \{\{1,2\},\{1,3\}\}$. The function $\{1,2\}$ maps any binary string $s = s_3s_2s_1$ into a new string by flipping values of the 1st bit s_1 and the 2nd bit s_2 simultaneously. Intuitively, applying this function will mean to change the decisions on p_1 and p_2 simultaneously. The function $\{1,3\}$ is similarly defined.

x	$M_2(x)$
x_0: fund none	111
x_1: fund p_1 only	110
x_2: fund p_2 only	101
x_3: fund p_1 and p_2	100
x_4: fund p_3 only	011
x_5: fund p_1 and p_3	010
x_6: fund p_2 and p_3	001
x_7: fund all three projects	000

The set of local maxima for Member 2 with (M_2, A_2) is $\{x_1, x_6\}$ which differs from Member 1's set of local maxima. The maximum of the value function x_6 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. x_1 is a local maximum for Member 2 for the following reasons: she encodes x_1 into 110 and x_1 has a value 40; applying trick $\{1,2\}$ gives 101 which corresponds to x_2 with value 20; applying trick $\{1,3\}$ to

110 leads to 011 which corresponds to x_4 with value 30; thus no improvement can be made through applying the two tricks that she has.

The above information tells us that if Member 1 and Member 2 work separately, either may end up with a solution that is not the best solution because their problem solving ability is confined by their respective perspective/heuristic pairs. Imagine they work as a team in the following way: 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 this team of Member 1 and Member 2 will always be able to locate the best solution, x_6. For example, suppose Member 1 works on the problem and gets stuck with x_5 which is one of her local maxima. Member 2 joins her and takes the solution x_5 located by Member 1. In Member 2's encoding, x_5 is 010 string. Member 2 applies trick $\{1, 2\}$ to get string 001 which corresponds to x_6 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 got stuck.

For a group of problem solvers, an object is a local maximum if and only if it is a local maximum for each problem solver within the group. To evaluate the expected performance of an individual or a group requires a precise 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.[8] Together with a prior probability distribution over objects, the Markov process generates a probability distribution over the set of local maxima allowing for the computation of the expected value of a local maximum for a problem solver. The expected value of a group of problem solvers can be calculated similarly.

The next section contains a special case of our general model in which all problem solvers rely on the same internal language – binary strings of length n. Although binary strings may seem an odd choice, they offer two methodological advantages: they are easy to understand and they map nicely into some economic problems such as Example 2 above. The restriction to a single internal language raises the spectre of brittleness. However, the abundant neighborhood structure in the binary encoding makes our results more powerful: many of our claims rely on the construction of problem solvers whose perspective/heuristic pairs combine in interesting ways; the fact that we can generate these perspectives in a more structured environment suggests that these findings hold more generally. The robustness of the findings is discussed in more detail at the end of this paper. Moreover, a primary goal of this paper is to present a coherent model of heterogeneous boundedly rational agents solving difficult problems. The binary string model enables us to contrast problem solving,

[8] See Ryall (1995) for a more complete Markov analysis.

information processing, service providing, and manufacturing firms.

4 A Binary String Model

Hereafter, the *objects* are binary strings of length n, denoted by $S = \{0, 1\}^n$, and all problem solvers use the same internal language. This internal language is also binary strings of length n. Note that this does not imply that all agents encode strings identically. This construction enables the introduction of an "identity perspective" in which a problem solver encodes the objects with the identity mapping. We often use the identity perspective as a benchmark in our analysis. 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 $V : S \to \Re$.

Binary strings map nicely into many economic problems. Letting 1 denote "yes" and 0 denote "no", a binary string can denote the set of potential 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.

Each problem solver has a *perspective*, which is her internal encoding of the objects as binary strings.

Def'n: *A* **perspective** $M : S \to S$ *is one–to–one and onto.*

In this definition, 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.

An algorithm consists of a finite set of instructions (Knuth 1968). Given that the number of strings is finite, an algorithm could list a different instruction for each string and still satisfy this definition. We rely on a more restrictive notion. We restrict attention to search rules defined independently of the string. For example, in the case of binary strings of length three, a heuristic might consist of the following three rules which generate, for each string, neighboring strings to be evaluated:

1. flip the first two bits
2. flip the first and last bits
3. flip all three bits

The neighbors of the string 000 would be 011, 101, and 111.[9] Similarly, the neighbors of the string 001 would be 010, 100, and 110. We formalize this idea as

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

follows: Define a *flipset* to be a subset of bits. A flipset can be viewed as a map from S onto itself. Let $N = \{1, ..., n\}$.

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

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

It is straightforward to show that the binary operation of composition in the set of flipsets is both associative and commutative. We define a class of flipset heuristics as follows. Let $m \geq 1$.

Def'n: *A* **flipset heuristic** $A = \{\phi_1, \phi_2, ..., \phi_m\}$.

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

Def'n: *The* **elementary heuristic** $A_E = \{\phi_1^e, \phi_2^e, ..., \phi_n^e\}$, *where* $\phi_i^e = \{i\}$

The elementary heuristic can also be defined for subsets of the set of bits. This becomes important when a heuristic does not apply to the entire domain, such as when a large problem is broken down into subproblems.

Def'n: *The* **elementary heuristic defined on** $K \subseteq N$, $A_E^K = \{\phi_i^e : i \in K\}$.

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.

A problem solver in this framework is defined by, $P = (M, A)$, where M is her perspective and A is her flipset heuristic. Following the general model in Section 3, we define:

Def'n: *The* **neighborhood of object** s *for* $P = (M, A)$, $b_M(A, s) = \{s' \in S : M(s') = \phi(M(s))$ *where* $\phi \in A\}$

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* $s \in S$, $b_{M_1}(A_1, s) = b_{M_2}(A_2, s)$.

Two problem solvers may have different perspectives and different heuristics but still are of no help to each other if the neighborhood structure generated by their respective perspective/heuristic pairs is the same. On the other hand, given any object, if the two problem solvers have very different neighboring objects, there is more chance that one problem solver can help to improve upon a local maximum of the other problem solver. Thus, we measure the diversity of two problem solvers by the average number of objects that belong to only one person's neighborhood.

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 some of these concepts with Example 2 in Section 3. The two city council members are denoted by $P_1 = (M_1, A_1)$ for Member 1 and $P_2 = (M_2, A_2)$ for Member 2. Using the binary string model, 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	$M_1(s)$	$b_{M_1}(A_1, s)$	$M_2(s)$	$b_{M_2}(A_2, s)$	$\cup_s \backslash \cap_s$
$x_0 : 0$	0	1,2,4,	7	3,5	5
$x_1 : 1$	1	0,3,5	6	2,4	5
$x_2 : 2$	2	0,3,6	5	1,7	5
$x_3 : 3$	3	1,2,7	4	0,6	5
$x_4 : 4$	4	0,5,6	3	7,1	5
$x_5 : 5$	5	1,4,7	2	0,6	5
$x_6 : 6$	6	2,4,7	1	5,3	5
$x_7 : 7$	7	3,5,6	0	4,2	5
$\Delta(P_1, P_2) =$					5

5 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. We derive three theorems that taken together establish the following two facts: First, in the binary string model, perspectives permit greater flexibility than heuristics; 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 equals 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 claim that follows uses the elementary heuristic as a benchmark to view equivalence.

Claim 1 *A perspective/heuristic pair (M, A) is equivalent to (M', A_E^K) for some perspective M' and some $K \subseteq \{1, 2, ..., n\}$ such that $|K| = m$ iff $A^{\oplus}$ is isomorphic to an m-dimensional hyperplane of S.*

Claim **??** implies that perspectives are at least as flexible as heuristics. Any problem solver, no matter the perspective and heuristic as long as the span of the heuristic has the full m-dimensions, can be looked upon as another problem solver with a possibly different perspective and a simple heuristic which is an elementary heuristic defined for a hyperplane of S.

Claim 2 *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.*

We argued earlier that problem solvers not only use different search rules but also represent problems differently. In our binary string model, Claims **??** and **??** together imply that differences in perspectives fundamentally enlarge the set of possible ways to solve a problem. If we believe that individuals represent problems differently, we believe then that the number of ways of solving a problem exceeds the number of different search rules.

Claim **??** suggests a lower bound on the number of equivalence classes of all problem solvers.

Claim 3 *The number of equivalence classes is at least*

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

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

6 Optimality Through Diversity

In this section, we discuss our results on diversity and optimality. The three claims 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 collective solutions. There are two possible interpretations. In the first, we imagine the problem as existing in the economy at large. It may be something concrete like designing a mousetrap or something abstract like teaching calculus. Each agent sees the current existing solution and attempts to improve upon it. If the agent makes an improvement, she earns the rents accruing from the increase in value. In this interpretation, there are no problems with incentives or communication. Recall that all agents have the same value function.

In the second interpretation, agents belong to a firm. Here the agents work together to solve a problem within their organization. Within the firm, there is a possibility of incentive problems, differences in values of solutions, and communication problems, since ideas need not be actualized. These caveats notwithstanding, our findings suggest that firms, by focusing diverse skills on common problems, can enable agents of limited abilities to locate optimal solutions to difficult problems. Thus, firms might be thought of as devices for coordinating the attention of agents.

We shall dichotomize the notion of diversity along the perspective and heuristic dimensions. Our first claim states that there exists a perspective such that the elementary heuristic has a unique local maximum, namely the global maximum. 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.

Claim 4 *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 maxima 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 $s \in L((M, A_E), V)$.

The result of Claim **??** suggests the possibility that a potential employee can locate an optimal solution to a difficult problem confronting a firm. Moreover, she need not be particularly advanced in the heuristic she employs, so long as she has the right perspective. A second implication, the one we shall stress here, is that diversity of perspectives can lead to location of the global optimum.

At first blush, this would appear to make the categorization of the difficulty of a function problematic: all functions are easy for someone. Yet, upon reflection, one recognizes that measuring the difficulty of a function depends upon the perspective of a problem solver.

We next consider the case where all agents rely on the same perspective but use different heuristics. We state two claims here. The first states that with enough diversity in heuristics, the optimal solution to a problem can always be guaranteed.

Claim 5 *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.*

pf: Let each agent's set of heuristic consist of a single flip set. Assume that there are $2^n - 1$ number of agents and each possesses a unique heuristic. The result follows.

The proof above is somewhat unsatisfying in that it relies on checking every possible solution. We now prove a stronger result which borrows results from Page (1995,1996). 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 hillclimbing (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 iff the elementary heuristic has a unique local optimum which is the global optimum. It has an ascent size of two iff 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 roughly as the appropriateness of a perspective. The smaller the ascent size, the better the perspective for the function.

Claim 6 *If a function V has an ascent size of k given the perspective M, then an upper bound on the minimal number of problem solvers with n flipsets and the common perspective M is given by*

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

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:

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

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

[10] Ascent size can also be measured relative to upper contour sets. Page (1995b) shows that as the function value improves, the ascent size weakly decreases.

The reason that m_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 claim is that as problems become harder more problems solvers are needed to guarantee locating the optimum.

7 Problem Solving Firms

In Section 2, we presented data showing an increasing number of people who solve problems as part of their work. A natural question to ask is whether constant, decreasing, or increasing marginal returns would hold for problem solving firms. Any insights into how the expected value of local optima located by a group of problem solvers varies with the number of problem solvers in the group would be of economic importance.

Although casual intuition suggests that problem solving would yield decreasing returns to additional problem solvers, several provocative insights emerge from our analysis which lead us to question such a conclusion. First, we 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. Second, we can construct value functions and problem solvers of equal ability who generate both increasing and decreasing returns to scale depending upon the order that they are applied to the problem. That two individuals have equal ability means that the expected value of local optima reached by either of the individuals alone is the same.

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. To generate their results, they manipulate income effects. We manipulate local optima and the basins of attraction of the various problem solvers to construct similarly counterintuitive results. Our results, though nonstandard, should not be interpreted negatively. Instead, they should be seen as suggestive that problem solving has less regularity than manufacturing. To generate decreasing returns to adding problem solvers, we derive one set of sufficient conditions. It requires that a priori none of the problem solvers has any insight into the problem at hand. Such an assumption runs counter to basic economic intuition. To remain in existence, problem solving firms must 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. Unfortunately, problem specific knowledge creates the possibility for interesting irregular phenomena such as those that we prove exist below: *anything can happen; order can determine the returns to additional workers.* Again, we want to be careful not to over interpret our findings. We feel that some confidence can be attached to the idea that problem solving firms have less regularity in their marginal products of labor. Whereas a producer of automobiles knows with great precision how many additional cars are produced if additional workers are hired, a software development firm or a movie studio, has less certainty in the value added from adding problem solvers.

7.1 Group Problem Solving

For problem solving firms, the goal is to locate solutions with the highest values. 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 local optima for the group. This performance measurement depends on the way we model how each individual searches for a solution and how a group works together to search for a solution. We first look at individuals. Each problem solver applies her flipsets sequentially. Any time she obtains an improvement, she updates the status quo.

Def'n: *Let $\sigma \in \Sigma(\{1, 2, ..., m\})$, where $\Sigma(\{1, 2, ..., m\})$ is the set of permutations on 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* $a = 1$

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.*

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

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

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 case 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 qualitatively for our purposes. The claims below hold for several alternative and equally plausible descriptions of group behavior. For mathematical convenience, we assume that a group of problem solvers act sequentially. The first problem solver applies her heuristic until reaching a local maximum. The second problem solver then joins the first problem solver. Together, the two act as a single problem solver. When that subgroup has located a local maximum, the next problem solver joins, and the three act as a single problem solver.

Let $\{P_j\}_{j=1}^i$ where $i > 1$ be a group of problem solvers. For each $j \in \{1, ..., i\}$, $P_j = (M_j, A_j)$ where $A_j = \{\phi_{j1}, \phi_{j2}, ..., \phi_{jm_j}\}$. For any $\{\phi_{h_1 g_{h_1}}, \phi_{h_2 g_{h_2}}, ..., \phi_{h_l g_{h_l}}\}$, a subset of $\cup_{j=1}^i A_j$, $h_a \in \{1, ..., i\}$ and $g_{h_a} \in \{1, ..., m_{h_a}\}$ for any $1 \leq a \leq l$,

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 exactly that 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.*

In the following definition, for any $1 \leq k \leq i$, $\sigma_k \in \Sigma(I_k)$ where $\Sigma(I_k)$ denotes the set of permutations 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.

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 in this way, we can compute the marginal value of adding problem solvers, in standard economic language, the marginal product of problem solvers (MPPS). 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. Formally we would have to write MPPS(i, J), where i is the problem solver, and J is the 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).

7.2 Returns to Additional Problem Solvers

Our next claim states that we can find a group of employees and a value function 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 maxima.

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. The order determines where the former local optima go. In one order, the former 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 former 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 to scale 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 very small but positive number. The strings with exactly three ones, of which there are twenty, play a prominent role in this example. Let s_k^3 where $k = 2$ to 21 denote these twenty strings.

We write problem solver 22's perspective as follows:

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

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

$$M_j(x_0) = 000000$$
$$M_j(x_j) = 111110$$

$M_j(x_{j+1}) = 111111$
$M_j(x_{\lfloor j+k \rfloor}) = s_k^3$ for $k = 2$ to 21 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 dir	x_1 $\rightarrow$	x_2 $\rightarrow$	x_3 $\rightarrow$	x_4 $\rightarrow$	x_5 $\rightarrow$	x_{10} $\rightarrow$	x_{11} $\rightarrow$	x_{20} $\rightarrow$	x_{21} $\rightarrow$	x_{22} $\rightarrow$	x_0 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 formal claim. The proof is constructive. Here the example differs in that the different problem solvers get rid of different numbers of local optima. Each former local optima is mapped into the global optimum.

Claim 7 *For any $m \geq 3$, there exist an $n \geq 1$, a group of problem solvers $\{P_i\}_{i=1}^m$, a value function V defined on binary strings of length n, 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 refer to the case $m = 5$ as a benchmark. 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 whose value is not equal to 1, and let s^{0j} for $j = 1$ to n denote the n strings with exactly one bit whose value is not 0. The key to the proof is to 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. Other local optima all have value -2. 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 problem solver and each group of problem solvers. For ease of exposition, we don't keep track of them specifically.

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 provide the basis for 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. Obviously $\delta > 0$. 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 immediately by construction.

We omit the proof of the claim for a general m as it follows the $m = 5$ case exactly.

The next claim states that for any nondecreasing 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.

Claim 8 *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.*

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:

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

The construction of the value function is very similar to that used in the proof of Claim **??**. The domain of the value function consists of 2^n object strings. As before, let s^0 denote the string of all 0's, s^1 denote the string of all 1's, s^{1i}'s for $i = 1, ..., n$ denote n strings with exactly one bit 0, and s^{0j}'s for $j = 1, ..., n$ denote n strings with exactly one bit 1. For each object string s, we define the value function $V : S \rightarrow \Re$ 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 we are going to define all 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 $i = 1, ..., n$, strings here are M_k-strings. By simple calculation, the probability of the search ending at a given s^{1i} is $\frac{1}{2n}$ 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, $E[V : \{P_i\}_{i=1}^{j}] = f(j-1) + \frac{\Sigma_j - \Sigma_{j-1}}{2n} \cdot 2 = f(j)$.

7.3 Maximally Difficult Functions

The previous results rely on constructive proofs. As we pointed out earlier, they demonstrate the possibility of irregularity in marginal products of labor in problem solving but they do not indicate the likelihood of such phenomena. They do not say 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 maximally difficult function.* A maximally difficult function has random values from the problem solver's perspective (Macken, Hagan and Perelson 1990). Equivalently,

Def'n: *A function V is* **maximally difficult** *for a problem solver $P = (M, A)$ if the problem solver's perspective is a randomly chosen one to one map from the binary strings of length n onto themselves.*

If a function is maximally difficult for a problem solver, then the problem solver has no insight or understanding of the problem.

Claim 9 *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 maxima is $1 - \frac{\theta}{m}$, where θ is some constant. (See Section 4 for the definition of a group of problem solvers being maximally diverse.)*

pf: See Macken, Hagan and Perelson (1990).

It follows immediately from the proof of the claim above 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. It does not imply that they always do better as evidenced by the following example.

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 easily been 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 maximum; (3) Group b (P_3 and P_4 together) can not always locate the global maximum, for example, the object string 2 (which has a value of 80) is a local maximum 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 attack 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. The problems of most relevance to our analysis should belong to the interesting "in between". Non difficult problems can be solved using calculus and maximally difficult problems can be solved using probability and measure theory. When a problem is difficult, but has some exploitable structures, mathematical models are more difficult to construct.

8 Discussion

In this paper we have constructed a model of economic agents of diverse problem solving behavior. The model distinguishes between perspectives, how people encode problems, and heuristics, how they go about trying to solve them. We have concentrated on a binary string model within this general framework. Using the binary string model, we show the benefit of explicitly modelling perspectives: adding perspectives significantly enlarges the possibility for diversity. Our finding – that diversity leads to optimality – should delight many economists. Collectively agents of limited ability can solve very difficult problems. We also show that the path to optimality may not possess much regularity vis a vis the returns. 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. Taken together these results suggest a lack of regularity in returns to adding problem solvers. Empirically, we may expect to see problem solving firms having greater variation in performance.

One obvious question to ask is what happens on average. Would not the local nonconvexities average out? Moreover, doesn't simple logic argue that eventually the returns to additional problem solvers diminish? The second question is valid but moot. By the time a firm can be certain that additional problem solvers will make no improvements, the firm has probably already hired too many. The first question requires a careful answer. Firms producing software, movies, consulting reports, drugs, and legal cases may not have many discrete outputs. Averaging may be over only one or two instances. If so, the implications of our findings stand. The implication being that if a firm solves big problems rarely, then anything could happen. If the firm solves many small problems, some averaging may occur. However, if the problems are similar, people obtain credit according to where they are in the hierarchy.

The interpretations so far have been limited to human problem solvers. Yet, the perspective/heuristic framework can equally well be applied to 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 modelling computers within our framework, a difference between humans and computers becomes evident. 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 immediately 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.[11]

A potential criticism of the theoretical results contained in this paper is that they are formulated in a binary string model and not in a more general framework. A detailed examination of the proofs supports the generalizability of these results. 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 are in fact quite general. In fact, we have proven them for arbitrary symmetric graphs. Binary strings happen to provide a convenient language for presenting the idea.

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, or they may be the result of different beliefs about an uncertain world. 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.

[11]If one encoding can be changed to another using a shortcut, then the analysis of this paper suggests that in fact the second encoding heuristic pair is probably equivalent to the first encoding with some other heuristic. Also, note that another difference between human and artificial problem solvers stems from the endogeneity of perspectives and heuristics. Humans may have almost no flexibility in their choice of perspectives and only limited range over their choice of heuristics.

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, Jennifer A., Jeffrey T. Polzer, Sigal Barsade, and Margeret Neale “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] Debreu, G., (1959). *A Theory of Value,* Wiley, New York.

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

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

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

[9] Hong, L. and S. Page (1998). “Diversity and Optimality” mimeo.

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

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

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

[13] Kirschenheiter, M.T. (1993). “Hard and Soft Information”, mimeo, Columbia University.

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

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

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

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

[18] 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.

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

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

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

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

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

[24] Page, S. (1995a). "Covers: A Contour Based Measure of Decomposability," *Complex Systems,* forthcoming.

[25] Page, S. (1995b). "Two Measures of Difficulty," *Economic Theory*, forthcoming.

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

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

[28] 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.

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

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

[31] Ryall, M. (1995). Personal communication.

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

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

[34] Thomas, David A and Robin J Ely "Making Differences Matter: A New Paradigm for Managing Diversity" *Harvard Business Review* September–October 1996 #96510

[35] 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

In this appendix, we shall prove the three claims in Section 5 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* $\oplus : \wp(\{1, 2, ..., m\}) \times \wp(\{1, 2, ..., m\}) \to \wp(\{1, 2, ..., m\})$ *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* $s \in S$.

Claim 1 *A perspective/heuristic pair* (M, A) *is equivalent to* (M', A_E^K) *for some perspective* M' *and some* $K \subseteq \{1, 2, ..., n\}$ *such that* $|K| = m$ *iff* $A^{\oplus}$ *is isomorphic to an* m*-dimensional hyperplane of* S.

To prove Claim **??**, 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* $K \subseteq \{1, 2, ..., n\}$ *such that* $|K| = m$ *iff* $A^{\oplus}$ *is isomorphic to an* m*-dimensional hyperplane of* S.

pf: ($\Rightarrow$) 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 $A_E^{K\oplus}(M_1(s^0))$ is an m-dimensional hyperplane of S.

($\Leftarrow$) When $A^{\oplus}$ is isomorphic to an m-dimensional hyperplane of S, 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 $K \subseteq \{1, 2, ..., n\}$ such that $|K| = m$, 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 $K \subseteq \{1, 2, ..., n\}$ such that $|K| = m$, 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. 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 $\sigma : \{1, 2, ..., m\} \to K$ such that $\phi_k(s) = M_1^{-1}(\phi_{\sigma(k)}^e(M_1(s)))$. 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 $\sigma : \{1, 2, ..., m\} \to K$ such that $M^{-1}(\phi_k(M(s'))) = M'^{-1}(\phi_{\sigma(k)}^e(M'(s)))$. This implies $\phi_k(s) = M_1^{-1}(\phi_{\sigma(k)}^e(M_1(s)))$ since $s' = M^{-1}(s)$ and $M_1 = M' \circ M^{-1}$.

Claim **??** is much simpler. It says

Claim 2 *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. Claim **??** 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, $1 \leq m \leq n$*$\}$

Claim 3 *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 Claim **??**, 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 **??**, 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 **??**: By Lemma **??**, 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 $(M', A_E^{K'})$.