DRUID

DRUID Working Paper No. 07-08

The Human Version of Moore-Shannon's Theorem:
The Design of Reliable Economic Systems

By

Michael Christensen and Thorbjørn Knudsen

Danish Research Unit for Industrial Dynamics

www.druid.dk

The Human version of Moore-Shannon's Theorem: The Design of Reliable Economic Systems

Michael Christensen and Thorbjørn Knudsen
Strategic Organization Design Unit (SOD)
Department of Marketing & Management
University of Southern Denmark
DK-5230 Odense M
Denmark
Corresponding e-mail: tok@sam.sdu.dk

Abstract:
Moore & Shannon's theorem is the cornerstone in reliability theory, but cannot be applied to human systems in its original form. A generalization to human systems would therefore be of considerable interest because the choice of organization structure can remedy reliability problems that notoriously plaque business operations, financial institutions, military intelligence and other human activities. Our main result is a proof that provides answers to the following three questions. Is it possible to design a reliable social organization from fallible human individuals? How many fallible human agents are required to build an economic system of a certain level of reliability? What is the best way to design an organization of two or more agents in order to minimize error? On the basis of constructive proofs, this paper provides answers to these questions and thus offers a method to analyze any form of decision making structure with respect to its reliability.

Key words: Organizational design; reliability theory; decision making; project selection
Jel codes:

The Human Version of Moore-Shannon's Theorem: The Design of Reliable Economic Systems

Michael Christensen

Thorbjørn Knudsen, tok@sam.sdu.dk
Strategic Organization Design Unit (SOD)
Department of Marketing & Management, University of Southern Denmark
DK-5230 Odense M, Denmark

Moore & Shannon's theorem is the cornerstone in reliability theory, but cannot be applied to human systems in its original form. A generalization to human systems would therefore be of considerable interest because the choice of organization structure can remedy reliability problems that notoriously plaque business operations, financial institutions, military intelligence and other human activities. Our main result is a proof that provides answers to the following three questions. Is it possible to design a reliable social organization from fallible human individuals? How many fallible human agents are required to build an economic system of a certain level of reliability? What is the best way to design an organization of two or more agents in order to minimize error? On the basis of constructive proofs, this paper provides answers to these questions and thus offers a method to analyze any form of decision making structure with respect to its reliability.

Key words: organizational design; reliability theory; decision making; project selection
History: Final revision – Prior versions 2002, 2004, 2005

1. Introduction

Economic and social organizations can be designed to remedy errors caused by fallible human agents, but their design can also compromise the joint outcome of individual decisions. This is an important problem for every business operation, and of critical importance in financial institutions and military intelligence.

It is a notable feature of human decision making that errors in judgement occur even if there are no conflicts of interest (Stiglitz 2002). Reliability research shows that judgement in social and economic organizations is shot through with failure (Reason 1990, 2000), a point that the contribution of human error in medicine (Bogner 1994, Edmondson 1996, Kohn et al. 1999) has brought home with force. Similarly, financial institutions absorb substantial losses from default on credit (Dahiya et al. 2003).

Social and economic organizations have an important role in reducing human error and its consequences. This insight, that the design of social and economic organizations can improve the judgemental accuracy of fallible human agents, originates in the work of von Neumann (1956) and Moore and Shannon (1956a,b). Their work on mathematical information theory reached the highly remarkable conclusion that it is possible to build a reliable system out of unreliable components. This insight is now a cornerstone in statistical theories of network reliability (Goldstine 1961, Lomnicki 1973, Lynn et al. 1998, Balakrishnan and Rao 2001).[1] Our work builds on classical reliability theory and important recent advances in related streams of literature. On this basis,

---

[1] Network reliability typically concerns the probabilistic modelling and analysis of the breakdown of systems that are built out of components that are subject to two kinds of failure (Lomnicki 1973). The classical example is a network of identical relays that operates if and only if k out of the n relays function. Such ideas inspired Marschak and Radner (1972), Hess (1983), and Sah and Stiglitz (1985, 1986, 1988) to examine the possibility of improving judgemental accuracy in economic systems that include fallible human decision makers.

we offer an analytical framework that provides guidelines for the design of a very broad range of decision making structures, including management teams, boards of directors, specialist functions such as credit evaluation systems in banks, and more.

During the last decades, a burst of activity has led to a number of papers that explore the possibility of building reliable economic systems out of fallible decision makers. A literature has explored the design of groups that improve judgemental accuracy when decision makers consider dichotomous choices in terms of accepting and rejecting alternatives. This literature includes Gradstein and Nitzan (1988), Ioannides (1987), Nitzan and Paroush (1985), Sah and Stiglitz (1985, 1986, 1988). A number of studies have considered committee decision making (Ben-Yashar and Nitzan 1997, Li et al. 2001, Sah and Stiglitz 1988, Shapley and Grofman 1984) and related legal applications (Klevorick et al. 1984, Grofman et al. 1983). Sequential decision making in simple stylized structures are studied by Koh (1992a,b, 1994a,b) and Sah and Stiglitz (1985, 1986, 1988). The literature is reviewed in Sah (1991). Promising new efforts include Gehrig et al. (2000), Koh (2003), Knudsen and Levinthal (2007), and Visser (2000). The main insight gained in all of these studies is that a broad range of evaluation structures have design properties that can alter, and sometimes improve, the judgemental accuracy of decision making structures.[2]

While this literature has pointed to the possibility of designing organizations whose structural properties can enhance the judgemental accuracy of fallible human agents, a number of issues have limited further advances. Most, if not all, of the models of economic evaluation structures are based on the simple stylized models of centralized and decentralized decision making introduced by Sah and Stiglitz (1986, 1988). These models have been useful in pointing to the importance of economic evaluation structures, but they ignore the complications of most empirical structures and almost all of the possible theoretical structures. The alternative models from statistical reliability theory are useful in the case of networks made of simple components (Lynn et al. 1998), but they are of little help if we want to analyze human agents. The basic problem here is that the cornerstone of reliability theory, the Moore-Shannon theorem, does not apply to human systems. At the present time, we do not have an answer to some rather important questions. Is it possible to build a reliable social organization from fallible human individuals? What is the best way to design a social organization, given a limited number of fallible employees?

Moore and Shannon's work on electric circuits (Moore and Shannon 1956a,b) would provide the tools for designing reliable organizations, i.e. organizations that are efficient in an information processing sense. Unfortunately, Moore & Shannon's methods cannot be applied in its original form to the design of economic systems because of a fundamental difference between electric components and human agents. Human agents have powers of deliberation that by far exceed those of electric components. Individual human agents have the power of saying "no" or "yes" to a proposed alternative, a fundamental property enabling the delegation of decision rights in human organizations. Lower level employees can, within limits, be empowered to say "yes" or "no" on behalf of the organization. This possibility is entirely absent in Moore & Shannon's work on electric circuits because it assumes that individual components do not have the power to say "no".

The Moore-Shannon theorem is the cornerstone in reliability theory and a generalization to human organizations would be of considerable interest. In broad strokes, a human organization is composed of members who are individually capable of accepting or rejecting alternatives. The

---

[2] A related literature considers issues of delay in organizations with limited information processing capacity (Bolton and Dewatripont 1994, Radner 1993, Zandt 1999). A broad literature in organization and management science has addressed important issues of contingent organizational design (Burton and Obel 1984, Donaldson 1995, Thompson 1967, Rivkin and Siggelkow 2003), but the relation between evaluation structures and judgemental accuracy has not been studied in a systematic way. The prior work in contingency theory that is closest related to ours is Burton and Obel (1984) who pointed to a similar general approach. Insights gained in earlier stages of the present work have been provided in a number of works, including Christensen and Knudsen (2002), and Knudsen and Levinthal (2007).

ability of the individual agent is captured by a screening function and agents can refer alternatives to other decision makers. The screening function captures human error in the sense that agents sometimes reject "good" alternatives and sometimes accept "bad" alternatives. Our main result is a proof that shows how the Moore and Shannon (1956a,b) theorem can be generalized to such human organizations. Our proof is constructive. It provides answers to the following three questions. Is it possible to design a reliable social organization from fallible human individuals? How many fallible human agents are required to build an economic system of a certain level of reliability? What is the best way to design an organization of two or more agents in order to minimize error?

The intuition is as follows. Reliability is improved when both omission and commission errors are reduced. Reduction of commission error (accepting projects that should be rejected) requires a conservative mode of decision making. This can be achieved by placing fallible individuals in a hierarchical organization. One example would be an organization where each vertical layer is empowered to reject a candidate for a vacant position. Hiring requires that all layers say "yes" and not even a single layer says "no". As the vertical chain of command is lengthened most bad proposals get weeded out along the way, but so do the good. Thus an infinitely long chain of command can eliminate commission error because it will, at the limit, not accept a single project. Unfortunately, this happens at the expense of throwing out all the good candidates. To counterbalance this effect, the organization must be flattened. A flat organization is one where employees are empowered to individually say "yes" and also refer rejected proposals to their neighbors for a second vetting. It seems possible that reliability of the organization could be improved by balancing vertical and the "flat" mode of organization. Our proof shows that this is possible and how it can be done. A practical application from credit evaluation in a bank illustrates how it can benefit practice.

The present article contributes in a number of ways to the body of knowledge on the optimal design of organizations. A new version of the main proofs and results of Moore and Shannon (1956a,b) are provided. Our main result is that perfectly reliable organizations can be built out of human agents that have powers of deliberation that are much greater than those of electric components. Thus, a second advantage is that we are able to provide a unifying method to design reliable economic systems on the basis of an adapted and enhanced version of (Moore and Shannon 1956a,b). A third advantage is that we provide results that are useful for the many complicated hybrid structures that we see in actual organizations, such as banks and military intelligence. A fourth advantage is that we are able to make a sharp distinction between the remedies that are required to remove judgemental bias in evaluation structures and the remedies that are required to increase the unbiased judgemental ability of an evaluation structure. We develop a method that can be used to achieve incremental improvements even for small economic systems. Finally, our analytical framework has the advantage that it can readily be applied to practice (as shown in our case study on credit evaluation).

The paper is organized as follows. In section 2 below, we specify the basic model including a very broad range of decision making structures spanned by the hierarchy and polyarchy of Sah and Stiglitz (1986). Section 3 introduces the extremity theorem (Theorem 1) which identifies hierarchies and polyarchies of any size, n, as evaluation structures that bound all other structures of size n with respect to screening. Together with Theorem 2, it provides a building block that is necessary to establish the main result of the present article. Two further results to be used in the main proof are established in section 3. The first result (Theorem 3) identifies the decision structures that are optimal with respect to increasing the judgemental discrimination for an organization comprising a given number of members. The second result (Theorem 4) identifies structures that effectively remove bias in an organization comprising a given number of members. These results provide the components for a constructive proof that make our analytical results readily available for practical applications. Our main result, the theorem of perfection (Theorem 5), is established in section 4. The proof of Theorem 5 uses the techniques of construction provided

by Moore and Shannon (1956a,b). Theorem 5 holds for human agents with powers to individually accept and reject alternatives on behalf of the economic system, and it is valid even in the case of non-monotone agent screening or non-monotone graph screening (Theorem 6). Theorem 5 is the human version of the Moore-Shannon Theorem on electrical components. A case study of credit evaluation in a bank illustrates how our analytical framework can be used in practice. Section 5 concludes the article.

2. The Model

Following the usual approach, we study economic systems whose members are decision makers (*agents*) that evaluate a set of alternatives (*projects*) and make a binary choice (*screening*). We refer to such economic systems as evaluation structures or decision making structures. The economic systems under study face the problem of choosing which projects to accept and which to reject. It is a notable feature of the model that acceptance is equivalent to the undertaking of a project. When a project is accepted it is impossible to avoid any of the economic consequences (whether good or bad); the economic system has made a commitment of unavoidable economic consequence.

There is stream of projects available that are independent and identically distributed (IID) as a random variable $\tilde{\mathbf{x}}$. Each project is represented by a vector of signals, $\mathbf{x}$. The vector of signals maps onto a scalar economic value, $P(\mathbf{x})$, a net income that is obtained by the economic system if the project is accepted. This scalar valuation is a measure of the true income of a project, including all of the relevant benefits and costs associated with the undertaking of a project.[3]

The task of the individual decision maker within the economic system is to evaluate projects. An evaluation is a binary choice, whether to accept or reject a project, made on the basis of the vector of signals about the project. The individual decision maker is characterized by an ability to evaluate projects. This ability is captured in the *agent screening* function, $f(\mathbf{x})$, which is a probability measure mapping each project onto a probability that the decision maker accepts the project. We assume that the abilities, $f(\mathbf{x})$, of the individual decision makers are statistically independent. Imperfect screening captures all sources of noise involved in making a decision. Our framework is robust as regards shifts in the source of noise, e.g. projects, signals about projects, communication channels, information processing technology.

An omniscient decision maker would not make a single error of judgement. Such a decision maker would process project signals without noise. The omniscient decision maker therefore accepts projects with economic value $P(\mathbf{x}) \geqslant 0 \, (< 0)$ with probability $f(\mathbf{x}) = 1 \, (0)$. That is to say, the agent screening function of the omniscient decision maker, $f_\Theta(\mathbf{x})$, is a composition of the Heaviside step-function and the economic value, $f_\Theta(\mathbf{x}) \equiv (\Theta \circ P)(\mathbf{x})$.

In actual practice, human decision makers are fallible; they make errors of judgement. Using the common analogy from statistical inference, a distinction is made between errors of commission and errors of omission. Decision makers accept projects that should be rejected (Type-II error) and they reject projects that should be accepted (Type-I error).

Noisy signals introduce error in judgement. The larger the noise, the less the ability to pass judgement in the sense of discriminating between projects with positive and negative economic value. In the case of maximal noise, the agent has no discriminating ability; the decision maker simply processes the signals about project quality by flipping a coin, $f(\mathbf{x}) = 1/2$. In the case of noiseless processing, the agent has perfect discriminating ability; the decision maker is omniscient

---

[3] The costs of making the decision is not included in the income because it is endogenous to the evaluation structure. The present effort abstracts from evaluation costs because it is focussed on the design of reliable economic systems in the sense that they minimize the incidence of Type-I and Type-II error subject to the constraints of the number of available evaluators and their ability.

and assigns signals about project quality to acceptance or rejection in a deterministic way. In the general case, the level of noise in the agent's processing of signals is the measure $\int_{\mathbf{x}} |f_{\Theta}(\mathbf{x}) - f(\mathbf{x})| \, d\mathbf{x}$. In particular cases, it is possible to make a sharp distinction between two forms of noise. One is judgemental bias, a deviation from symmetry in agent screening, and the other is imperfect but unbiased agent screening (Ben-Yashar and Nitzan 1997). In the following such a distinction is introduced at the level of evaluation structures.

There are n members in an evaluation structure. The task of the evaluation structure is to decide which projects to accept and which to reject. Its objective is to minimize the incidence of Type-I and Type-II error. The economic system aims to minimize error subject to the constraints of decision making ability and possibly system size.[4] We show how this problem can be solved with respect to internal structure and system size as choice variables, holding constant the decision making ability.

Consider a system of fallible agents that are *homogenous* in their decision making ability.[5] The sources of error is noise in the processing of signals about project quality. Each individual evaluator has access to two distinct types of communication channels, one is used in the case that a project is accepted and the other in case of rejection. It is the availability of both of these channels of communication that allow the evaluators to make the independent deliberate choices that are characteristic of human agents. It is possible for an evaluator to individually accept or reject a project on behalf of the economic system without interference by other members. The human agent can further choose the mode of transmission and the range of the possible receivers. Such powers of deliberation by far exceed those of the electric components that we find in Moore and Shannon (1956a,b)[6] and in the statistical theories of network reliability (Lynn et al. 1998). It is the actual design of the internal structure that determines the level of delegation. Decentralized structures have a high level of delegation and centralized structures a low level. Our framework applies for a very broad range of decision making structures spanned by the hierarchy and polyarchy of Sah and Stiglitz (1986), including organizations that can be modelled as deterministic, stochastic, or cyclical graphs.

The economic system is modelled as a graph. Each node represents a decision maker and each edge represents a channel of communication. We study homogenous graphs (one type of agent, A) with two types of edges (accept/reject). The entry and the exit of the projects are determined by the way the internal structure is connected to three external nodes: (1) the *initial portfolio* ($\mathbf{I}$) containing the distribution of projects $\tilde{\mathbf{x}}$, (2) the *final portfolio* ($\mathbf{F}$) where the accepted projects are implemented, and (3) the *termination* node ($\mathbf{T}$) where the rejected projects are dumped. The design of the evaluation structure involves the specification of the edges that connect members, the specification of the edges that connect the internal structure, through some of its members, to the external nodes ($\mathbf{I}, \mathbf{F}, \mathbf{T}$), and the specification of the rules that determine how many times a decision maker can evaluate the same project (a truncation rule).

The generalization of the agent screening function f to the level of a specific architecture G is the *graph screening* function F_G. The graph screening function is an aggregation rule that assigns individual decisions of acceptance and rejection to any structure. It can be viewed as a generalization of the aggregation rules that have previously been used to model decision making in the

---

[4] In the case of perfect decision making ability, the economic system has no effect because there is not made a single error; all projects with positive income would be accepted and all projects with negative income rejected.

[5] The e-companion extends some of the components of the current framework to heterogeneity in screening ability. Further extensions are possible, but would require treatment in a different article.

[6] Generally, Moore & Shannon's components are relays that are incapable of making individual decisions whether to "accept" or "reject" an eletric current on behalf of the entire system. The components of statistical theories of network reliability have similar limited powers of deliberation.

case of committees (Ben-Yashar and Nitzan 1997, Sah and Stiglitz 1988). In mathematical reliability theory, the graph screening function is know as the system reliability (Lomnicki 1973) or the reliability function of a network (Carlsson and Grenander 1966). The graph screening function, as well as other organizational properties, can be extracted through path enumeration, recursive expansion or similar schemes.[7]

It is often useful to separate the organizational screening capabilities from those of the agents. The pure organizational property is the *reduced* graph screening function which is a function of the project appearence, $\alpha \equiv f(\mathbf{x})$. In the case of homogenous ability, the reduced graph screening function is a polynomial in α, commonly known as the reliability polynomial.[8] Define the maximal (minimal) evaluation count as the maximal (minimal) number of evaluations that is available for the economic system before a decision must be made. These quantities bound the order of the graph screening polynomial. The highest order power of α in the graph screening polynomial is no higher than the maximal evaluation count and the lowest order power is no lower than the minimal evaluation count. Obviously, when decision makers are only allowed to evaluate a project once, the size of a decision making structure cannot be lower than the maximal evaluation count.

3. Fundamental Organization Structures

A few organization structures have fundamental properties as regards the improvement in reliability to be achieved by their application. The theorem of Moore and Shannon (1956a,b) approaches perfection by first removing bias in the imperfect screen, and then by steepening the imperfect screening, thereby improving the discriminating ability. This is done using so called *ladder* and *hammock* graphs.

The human version of the Moore-Shannon perfection theorem likewise relies on a small set of special graphs with remarkable screening properties. Sah and Stiglitz (1985, 1986) identified two archtypical evaluation structures: hierarchies and polyarchies. These are used here to handle special cases and reused in the following extension of the theorem. Sah and Stiglitz (1988) later identified these two structures as the extremes of a more generic set of structures: committees. The steepening is obtained by graphs related to selfdual committees, and they play the role of the *hammock* graphs in the original proof. The notion of a selfdual graph refers to a decision making structure that will induce a symmetrical graph screening function, i.e. a symmetrical reduction of Type-I and II errors. Finally, a new set of graphs, called *stair* graphs, are introduced which facilitate effective removal of bias in imperfect screenings, thereby playing the role of the *ladder* graphs in the original proof.

The theorems of the present article are proved in the e-companion, and all the proofs are constructive in order to not only guarantee existence of almost perfect organizations but to show exactly how such organizations are obtained. These construction methods are directly applicable in real business cases as is illustrated in section 4.4. The theorems of this section therefore serve two purposes. They are building blocks for the main perfection theorem and its extension (presented in section 4). And they are guides on how to improve discriminating ability and remove bias (illustrated in section 4.4).

---

[7] Christensen and Knudsen (2002) provide a procedure, which can extract the functions for any decision making structure, including structures with heterogenous members.

[8] In the case of j levels of heterogeneous ability, $\alpha_j \equiv f_j(\mathbf{x})$, the graph screening function is a multinomial in α_j.

3.1. Extremity of Polyarchies and Hierarchies

This section introduces a theorem that identifies hierarchies and polyarchies of any size, n, as evaluation structures that bound all other structures of size n with respect to screening. If the economic system is organized as a polyarchy, an acceptance of a project by any individual implies that the project is also accepted by the economic system. By contrast, in a hierarchy, a rejection of a project by any individual implies that the project is also rejected by the economic system.

The extremity theorem provides a necessary building block that enable us to provide a general tool for the design of reliable economic systems (section 4). A simple version of the theorem is presented here. The proof will be given in the e-companion to this paper, section EC.1, together with a number of extensions.

THEOREM 1. *For any positive integer n, let $\mathcal{G}_n$ be the set of all graphs that can be constructed from agents indistinguishable with respect to project screening and under common structural and dynamic requirements, such that the maximal evaluation count is no more than n. Let $P_n \in \mathcal{G}_n$ be the polyarchy and $H_n \in \mathcal{G}_n$ be the hierarchy of n members, having graph screening functions F_{P_n} and F_{H_n}, respectively. Any graph $G \in \mathcal{G}_n$ with graph screening F_G satisfies*

$$F_{H_n}(\mathbf{x}) \leqslant F_G(\mathbf{x}) \leqslant F_{P_n}(\mathbf{x}) \tag{1}$$

for all $\mathbf{x}$.

In extreme situations, when there are only projects with positive ($P(\mathbf{x}) > 0$) or negative income ($P(\mathbf{x}) < 0$), the n member polyarchy (hierarchy) will dominate any other structure. Provided the costs of making the decision in an evaluation structure are not prohibitive, the n member polyarchy (hierarchy) will also dominate the individual agent.[9]. Even if this result is rather trivial, it holds with remarkable generality (see the e-companion EC.1)

According to the extremity theorem, finite hierarchies (polyarchies) map the agent screening of any project closer to 0 (1) than any other structure with the same number of agents (provided the agents are homogenous in ability). The implications are that polyarchies have the smallest possible Type-I error (rejecting a good project) and the largest possible Type-II error (accepting a bad project), while hierarchies are the extremes in the exact opposite way. Both kinds of structures tend to reduce only one form of error. The minimal number of agents that are required to reduce the incidence of Type-I and Type-II error to some minimal desired level is provided below in Theorem 2. Section 4, then, shows how economic systems that include both hierarchical and polyarchical elements can be designed to minimize both Type-I and Type-II error.

THEOREM 2. *Given any threshold $0 < \delta < 1$ and a point $\alpha_0 \in]0, 1[$, the number of agents n in a hierarchy such that $F_{H_n}(\alpha) \leqslant \delta, \forall \alpha \in [0, \alpha_0]$, is*

$$n \geqslant \frac{\log(\delta)}{\log(\alpha_0)} \tag{2}$$

and the number of agents n in a polyarchy such that $F_{P_n}(\alpha) \geqslant 1 - \delta, \forall \alpha \in [\alpha_0, 1]$, is

$$n \geqslant \frac{\log(\delta)}{\log(1 - \alpha_0)} \tag{3}$$

---

[9] In situations with more realistic project distributions that include both positive and negative income, the optimal evaluation structures are hybrids that include both hierarchies and polyarchies.

3.2.   Optimality of Selfdual Committees

We now proceed to characterize the decision structures that are optimal with respect to increasing the judgemental discrimination for an organization comprising a given number of members. The aim is to identify the most effective symmetrical graph screening function, i.e. the decision making structure that most effectively reduces both Type-I and II errors. Such a symmetrical screening function is known as a selfdual graph. At the limit, a selfdual graph will approach a step function, i.e. projects below some criterion are accepted with probability zero and projects above this criterion will be accepted with probability one. That is, we identify the optimal selfdual graph:

THEOREM 3. *For any positive integer n, let $\mathcal{D}_n$ be the set of all selfdual graphs that can be constructed from agents indistinguishable with respect to project screening and under common structural and dynamic requirements, such that the maximal evaluation count is no more than n. The slope of the screening polynomials at $\alpha = 1/2$ of the graphs in $\mathcal{D}_n$ cannot exceed that of*

$$F_{G_{n*}}(\alpha) = \sum_{i=\frac{n+1}{2}}^{n} B_i \alpha^i \quad with \quad B_i = \frac{\Gamma(n+1)}{(\Gamma(\frac{n+1}{2}))^2} \frac{(-1)^{i-\frac{n+1}{2}}}{i} \begin{pmatrix} \frac{n-1}{2} \\ i - \frac{n+1}{2} \end{pmatrix} \tag{4}$$

and at least one graph $G_{n*} \in \mathcal{D}_n$ has this reduced graph screening.

Here, Γ is the Gamma-function, which reduces to factorial for natural numbers, $\Gamma(n+1) = n!$, and the stacking in the parantheses denote a binomial coefficient. Interestingly, the screening polynomial given by equation (4) matches the screening of a selfdual committee (Sah and Stiglitz 1988). Such a committee always consist of odd n members and consensus $k = (n+1)/2$, i.e. the simple majority rule. The proof in the e-companion shows how to build a graph with the optimally discriminating screening without the use of consensus rules. These are the graphs with the steepest graph screenings

$$F'_{G_{n*}}(1/2) = \frac{1}{2^{n-1}} \frac{\Gamma(n+1)}{(\Gamma(\frac{n+1}{2}))^2} \tag{5}$$

in the middle of the reduced screening interval.

3.3.   Stair Graphs

The final family of graphs used in the main proof is the so called *stair graphs*. A graph is a stair graph if and only if it has one entry point, no loops, and each agent has at most one channel of communication to another agent. The generic stair graph is therefore just a chain of agents arranged as a hierarchy whose last agent rejects to a polyarchy whose last agent accepts to a hierarchy etc. Like in the family of committees of various degrees of consensus, hierarchies and polyarchies are also the extremes of the family of stair graphs. In analogy with real stairs and assuming rejection is downward, the hierarchy is the vertically flat structure with no steps (floor; the project must be pushed all the way through), and the polyarchy is the horizontally flat structure (wall; the project easily falls over the edge).

The stair graphs have been chosen for (at least) three reasons. First of all they have monotonous graph screening polynomials (due to the fact that there is no branching in the structure), which is a crucial property if any of the work of Moore and Shannon (1956a,b) is to be reused in the current setup. Second, the size equals the maximal evaluation count. Third, they are very effective at removing bias as the following theorem shows:

THEOREM 4. *For any $0 < \alpha_0 < 1$ and any $0 < \varepsilon < d \equiv \min(\alpha_0, 1 - \alpha_0)$, a stair graph G exists with no more than*

$$n = \lceil \log(\varepsilon/2)/\log(d) \rceil \tag{6}$$

agents satisfying the relation:

$$F_G(\alpha_0 - \varepsilon) < 1/2 < F_G(\alpha_0 + \varepsilon) \tag{7}$$

The above theorem shows that any screening function can be modified via a stair graph to cross from mainly rejecting to mainly accepting projects at any reduced project value $\alpha_0 \equiv f(\mathbf{x}_0)$ with an arbitrarily low tolerance ε. Picking the shift point to occur within the set of projects of zero value, $P(\mathbf{x}_0) = 0$, will remove any bias of the original screening under the assumptions of perfect and positive correlation between screening and project value.

4.   Approaching Perfection

The perfect graph screening function, $F_\Theta(\mathbf{x})$, is identical to the perfect agent screening function, $f_\Theta(\mathbf{x})$. Moore and Shannon (1956a,b) showed (see theorem 6) that it is possible to build a graph out of unreliable electric components (relays) with a screening performance that deviates arbitrarily little from perfection. Their result built on the following assumptions: (1) the agents are able to discriminate between projects with positive and negative value, (2) the agents are more likely to accept projects with positive value than projects with negative value, i.e., the correlation between project value and agent screening is perfect and positive, and (3) the graph screening function is monotone (see Moore and Shannon 1956a, eq. (4)).

In the following we establish two results under the very general assumption that agents do not evaluate projects in a deterministic way, i.e. we drop assumptions (1)-(3) of the original proof. First, we establish the general result that social and organizations can be designed to approach perfection when a single criterion separates outcomes. Under assumption (1) and (2) this result becomes the human Moore-Shannon theorem. In that case, organizations can be designed to approach perfection when a single criterion separates *favorable* and *unfavorable* outcomes.

Second, we further widen the scope of our general result and show that social and economic organizations can be designed to approach perfection when multiple criteria separate outcomes. Under assumption (1), this becomes an extended version of human Moore-Shannon theorem, according to which organizations can be designed to approach perfection when multiple criteria separate *favorable* and *unfavorable* outcomes. Remarkably, our extended version of the Moore-Shannon theorem drops assumption (2).

4.1.   Perfection: The Single-Step Function

We now present the Moore and Shannon (1956a,b) result for economic systems. The projects can be divided into two disjoint sets according to the sign of their value. A perfect positive correlation between screening outcomes and project value ensures that the corresponding sets of screening outcomes are disjoint and ordered in a similar way. Consider the graph screening F as a function of the agent screening, $\alpha = f(\mathbf{x})$. Throughout, we refer to this as the reduced graph screening space and $F(\alpha)$ as the reduced graph screening function. As a consequence, the task of constructing a graph screening that deviates arbitrarily little from perfection can be described as constructing a graph whose reduced screening function shifts from a probability arbitrarily close to 0 to a probability arbitrarily close to 1 within an arbitrarily narrow interval around some point α_0. The point of the desired shift, α_0, in the graph screening function is any point that separates the two disjoint sets relating to screening outcomes.

The deviation from perfection can be expressed by two parameters ε and δ. As illustrated in Figure 1, the parameter ε defines the interval around α_0. The parameter δ defines the deviation
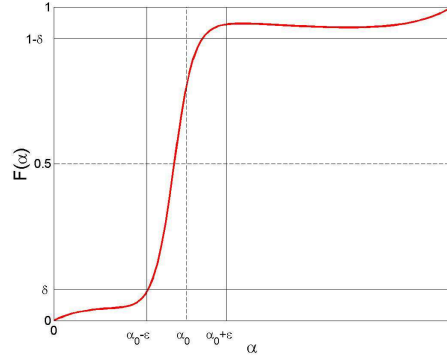
Figure 1 **A graph screening function that separates disjoint sets of project value on the basis of screening outcomes in a point α_0 satisfying condition (8).**

from the extreme screening outcomes of zero and one. The error-rates of the screening function should not exceed δ outside the interval $[\alpha_0 - \varepsilon, \alpha_0 + \varepsilon]$:

$$F_{G_n}(\alpha_0 - \varepsilon) < \delta \quad \wedge \quad F_{G_n}(\alpha_0 + \varepsilon) > 1 - \delta \tag{8}$$

The boundary points, $\alpha_0 \in \{0,1\}$, are trivially dealt with by polyarchies and hierarchies of increasing size, as shown in the extremity theorem (see Theorem 2). For any other α_0, we now prove a theorem that adapts the Moore and Shannon (1956a,b) perfection theorem to organizations of fallible human agents.

THEOREM 5. *Given any position $0 < \alpha_0 < 1$ for the shift in graph screening, any threshold $0 < \delta < 1/2$, and any radius $0 < \varepsilon < \min(\alpha_0, 1 - \alpha_0)$, then an architecture can be constructed from no more than*

$$n_{ck} \leqslant 25 \cdot \left\lceil \frac{\log(\varepsilon/2)}{\log(\min(\alpha_0, 1 - \alpha_0))} \right\rceil \cdot \left(\frac{1}{2\varepsilon} \right)^{\frac{\log(5)}{\log(11/8)}} \cdot \left(\frac{\log(3\delta)}{\log(3/4)} \right)^{\frac{\log(5)}{\log(2)}} \tag{9}$$

agents whose graph screening function, F_{G_n}, fulfills condition (8).

If $\delta > 1/3$ the number of agents to achieve the required level of reliability in Theorem 5 reduces to:

$$n_{ck} \leqslant 5 \cdot \left\lceil \frac{\log(\varepsilon/2)}{\log(\min(\alpha_0, 1 - \alpha_0))} \right\rceil \cdot \left(\frac{1}{2\varepsilon} \right)^{\frac{\log(5)}{\log(11/8)}} \tag{10}$$

The major obstacle in the generalization of the original proof (Moore and Shannon 1956a,b) originates in the possible delegation of decision rights. Electric currents simultaneously try out any possible path through the network. Even if one path is closed by rejection, any remaining open path will be checked. Not so with economic and social systems where a single agent may reject a project on behalf of the economic system even though other parallel paths through the architecture exist, a fact that often results in non-monotonous reduced graph screening functions. As a consequence, the fundamental equation (4) of Moore and Shannon (1956a), which is essential to the original proof, does not hold for all social and economic systems. Moreover, the specific graphs used in the original proof cannot be applied here as they are not general enough to include social agents that have the powers to reject or accept a project on behalf of the economic system. In order to overcome these obstacles, we extend Moore & Shannon's proof by including entirely new structures whose members have powers to reject or accept projects on behalf of the system.

The proof can be found in EC.2 of the e-companion, and it uses a technique similar to that of Moore and Shannon (1956a,b) for constructing an explicit graph. As in Moore and Shannon

(1956a,b), the proof will, for reasons of mathematical convenience, be provided in a construction process with three steps, referred to as the *opening game*, *middle game* and *end game*. The opening game consists of finding an architecture (a stair graph of Theorem 4) with n members such that the point, α_0, is the desired shift in the graph screening function, F_{G_n}, separating disjoint sets of positive and negative project value. That is to say, the graph screening function is moved over such that it intersects the value $1/2$ near α_0, $F_{G_n}(\alpha_0 - \varepsilon) < 1/2 \wedge F_{G_n}(\alpha_0 + \varepsilon) > 1/2$. The middle game consists in steepening the graph by recursive expansion (thus increasing n) such that $F_{G_n}(\alpha_0 - \varepsilon) < 1/4 \wedge F_{G_n}(\alpha_0 + \varepsilon) > 3/4$. Recursive expansion is the procedure of replacing each agent of a highly discriminating structure (a selfdual committee of Theorem 3) with a copy of the architecture found in the opening game. This procedure is known as composition. The endgame then consists in further steepening the graph by recursive expansion in order to obtain the required level of reliability with a total of n agents such that condition (8) is fulfilled.

4.2. Perfection: The Multi-Step Function

Theorem 5 shows that economic systems can be designed to deviate arbitrarily little from the limit of perfect reliability. This result is now established in the more general case. We require the minimal assumption that agents are able to discriminate between disjoint sets of positive and negative project value.

We do not require that the correlation between project value and agent screening is perfect, positive or even different from zero. That is to say, we do not require a monotone *agent* screening function which is an advantage because we are able to design economic systems that can repair serious human error.

Second, we do not require a monotone *graph* screening function. When we replace relays by human agents that have powers to individually accept or reject projects on behalf of the economic system, the graph screening polynomial is not necessarily monotone even if the agent screening is monotone. To the best of our knowledge this problem has not been recognized in previous research. In the following, we establish a new version of the Moore and Shannon (1956a,b) theorem of perfection that is valid even in the case of non-monotone agent screening or non-monotone graph screening.

Assume that agents are merely able to discriminate between projects with positive and negative value. That is, let projects fall within disjoint sets: $A_- \cap A_+ = \emptyset$. Here it will be assumed that the agent screening function and the value mapping are both piecewise continuous, which imply that $A_- \cup A_+$ is at most split into a finite number of intervals.

THEOREM 6. *Given any threshold $0 < \delta < 1$, a series of m (odd) shift points in reduced space*

$$0 < \alpha_1 < \alpha_2 < \cdots < \alpha_m < 1 \tag{11}$$

and a radius $0 < \varepsilon < \min_i(\alpha_{i+1} - \alpha_i)$, a graph can be constructed whose screening will jump from below δ to above $1 - \delta$ (and back alternatingly) within ε of the α_i's.

A graph displaying the desired multi-step screening function is obtained by first finding single-step graphs using a stricter $\delta' < \delta$ but reusing ε. Then these graphs are arranged into a hierarchy with every other subgraph rejecting to a suitably large polyarchy as illustrated in Figure 2 (3 jumps). An example of a multi-step graph screening function obtained from this procedure is shown in Figure 3 (5 jumps).
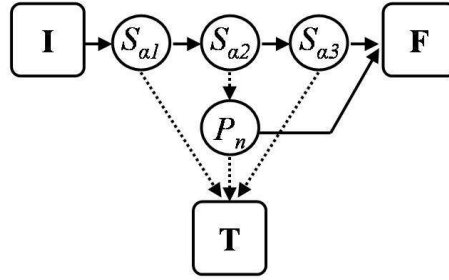
Figure 2   **Social and economic systems may have non-monotonous screening capabilities as function of project appearance, which allow multi-step functions shifting at $[\alpha_1, \alpha_2, \alpha_3, ...]$ (an odd number of shifts). These properties can be utilized to create more general architectures which screen perfect whenever the appearance of good and bad projects fall within disjoint sets. Full lines are acceptance edges and dashed lines are rejection edges.**

4.3.   Comparison with Moore & Shannon's Bounds

For comparison the original bound of Moore & Shannon translated to the present setup is:

$$n_{ms} \leqslant 81 \cdot \left\lceil \frac{\log(\varepsilon/2)}{\log(\min(\alpha_0, 1 - \alpha_0))} \right\rceil \cdot \left(\frac{1}{2\varepsilon} \right)^{\frac{\log(9)}{\log(3/2)}} \cdot \left(\frac{\log(\sqrt{8}\delta)}{\log(1/\sqrt{2})} \right)^{\frac{\log(9)}{\log(3)}} \tag{12}$$
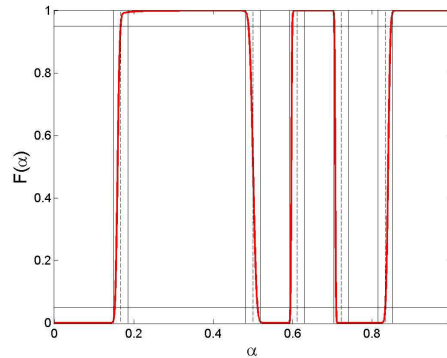

Figure 3   **Example of a multi-step function with $\delta = 1/20$, $\varepsilon = 1/54$ and shifts $[3/18, 9/18, 11/18, 13/18, 15/18]$ as constructed by the method devised in the proof of Theorem 5 and 6.**
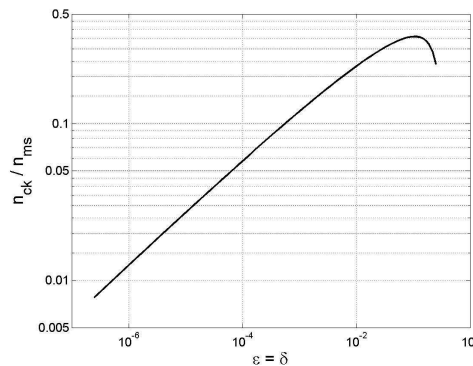

Figure 4   **The bound on the number of agents required to reach a given level of reliability found in the present study (n_{ck}) compared to the Moore and Shannon (1956a,b) bound (n_{ms}) for $\varepsilon = \delta$.**

Compared to the bound provided in the present article (equation (9)), this bound is slightly worse with respect to ε and slightly better with respect to δ. Since the present theorem applies to a wider family of architectures, this indicates that the dependence of ε is too pessimistic and the dependence of δ is quite realistic. This conjecture is readily confirmed by explicit construction.

4.4. An Example: Organizing Credit Evaluation in Bank F

We here illustrate how our framework can be used in practical applications. Our example is based on an actual case study of credit evaluation in a bank. Within the present context, we highlight some features of this case study.

Bank F has a number of local branches where Credit Advisors evaluate applications from business clientele. The evaluation results in immediate approval, rejection or, as is often the case, referrel to a Credit Officer in the bank's Central Credit Unit. The Credit Officer can approve, reject, refer the project to the next layer, or in some cases consult with a colleague at the same level (CAB). A number of advicers work at the CAB level, and one of these may finally approve or reject the project. Or, the project is pushed on to its final station, the Head of the Central Credit Unit. We are here considering applications of modest size (approximately \$1 million) that occur rather frequently (we considered a sample of 209 recent credit applications).

The common measure of the efficacy of a bank's credit evaluation is the number of defaults, a term that refers to the frequency of losses (error rates). The bank has a good estimate of Type-II error (defaults), but little information on Type-I error (rejecting good applications). The error rate for this bank is approximately 0.5% and there are significant differences among comparable banks. The official policy of this bank is to "caution all evaluators to be mindful of the balance between risk and reward." In practice, this translates into a conservative policy of evaluating a number of indicators that are thought to correlate with risk of default.

The bank's objective is to minimize the incidence of Type-I and Type-II errors subject to the constraints of the number of available evaluators (system size) and their ability. The probability of a Type-I error (rejecting a good project) is[10]

$$P_I = \int \Theta(P(\mathbf{x})) \left(1 - F_G(\mathbf{x})\right) \tilde{\mathbf{x}} \, d\mathbf{x} \tag{13}$$

where $\tilde{\mathbf{x}}$ is the distribution of projects, and the probability of a Type-II error (accepting a bad project) is

$$P_{II} = \int \left(1 - \Theta(P(\mathbf{x}))\right) F_G(\mathbf{x}) \tilde{\mathbf{x}} \, d\mathbf{x} \tag{14}$$

In order to derive specific prescriptions it is useful to consider the agent's discriminating ability against the system's tolerance of error, i.e. what is the tolerance for uncovered losses on a credit application? The system's tolerance of error is the maximal absolute project value for which errors have consequences that can be ignored, i.e. a relatively small loss. A system that tolerates no error whatsoever requires a zone of uncertainty that is zero. Such a system requires employment of an omniscient decision maker, $f_\Theta(\mathbf{x})$, but these are not on the job market. In our bank, the system must be designed so the tolerance of error includes credit applications with acceptable losses, but excludes the credit applications with unacceptably high losses.

From a design perspective, it is further useful to make a distinction between judgemental bias and ability. Judgemental bias is a deviation from symmetric screening around the point of zero project value. An employee may systematically be overoptimistic and accept credit applications

---

[10] Percentages (conditional probabilities) are readily obtained from these quantities and a performance measure can be constructed by a suitable weighting of the errors.
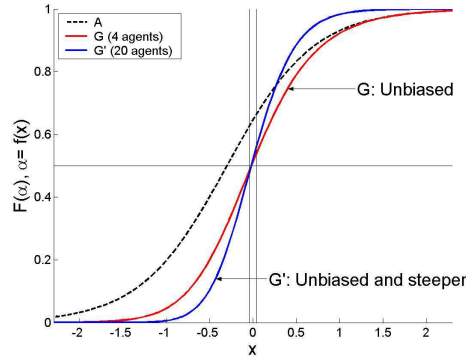
Figure 5   **Given a low tolerance $x_{tol} = 1/25$, a fairly large bias ($x_0 = -0.3$) can be removed with just 4 agents organized in a stair graph (a 2-member polyarchy in which the last agent accepts to another 2-member polyarchy). Using the procedure outlined in the proof, the graph can be further steepened if the structure is designed from 20 agents using one composition (can further be reduced to 12 agents).**

that turn out to be loosing propositions. Or, what is more likely, an employee may have a "healthy" conservative bias that tends to exclude good applications. A biased agent screening function is skewed. If the skew is significant, the system's tolerance of error may no longer include the agent's zone of uncertainty. The agent's judgemental ability is a different matter. Even an unbiased agent may have too little discriminating ability, which means that the agent's zone of uncertainty exceeds the system's tolerance for error. There are two complementary approaches to improving the system. The first is simply to hire evaluators with superior ability, and replacing evaluators with intolerable performance. The second concerns the design of the evaluation system given the present level of judgemental ability.

To illustrate application of our theory, we consider how bank F could achieve dramatic improvement of credit evaluation by choosing a design that both removes judgemental bias and increases the discriminating ability. We actually conducted an experiment in bank F in order to extract a screening function from 40 randomly selected credit evaluators. A mixture of twelve indicators that the bank commonly uses to evaluate credit applications were selected and a number of fake applications were constructed. The fake applications had known quality and frequency so it was possible to extract the average screening function.[11]

In the following example, we model credit applications as the scalar value, $P(x) = x$. We use a generic sigmoid agent screening function that is consistent with the empirical screening function in bank F:

$$f(x) = \frac{1 + \tanh\left(\frac{x - x_0}{\Delta x}\right)}{2} \tag{15}$$

where x_0 is an arbitrary bias and $1/f'(x_0) = 2\Delta x$ is the zone of uncertainty. The tolerance x_{tol} defines the interval $[-x_{tol}/2, x_{tol}/2]$ around zero containing credit applications of little consequence (most of the risk is covered by collateral and other securities). We set the scale of the zone of uncertainty to unity, $\Delta x = 1$, and measure the system's tolerance of error on this scale. That is to say, the agent's zone of uncertainty coincides with the system's tolerance when $x_{tol} = 2$. If the system's tolerance is significantly lower than the agent's zone of uncertainty ($x_{tol} < 1/10$), the agent is not sufficiently reliable. Note, the agent's screening function (including judgemental bias and ability) as well as the system's tolerance of error carries over into the reduced space because $\alpha_0 = f(0)$ and $\varepsilon/x_{tol} \approx f'(0)$. As the system approaches perfect reliability, $\varepsilon \to 1/2$, $x_{tol} \to 0$ and $f'(0) \to \infty$.

---

[11] It was sigmoid and it was fitted to a sigmoid tanh function with less than 0.5% unexplained variance.

Table 1   The number of agents needed in the stair graph (from the opening game) in order to remove any initial bias ($\alpha_0 \neq 1/2$) for any agent screening. The numbers in parantheses are polyarchies. Values of $\alpha_0 > 1/2$ are obtained by symmetry through dual graphs.

$\alpha_0 \setminus \varepsilon$	0.0001	0.0010	0.0100	0.1000	0.4000
0.45	4	4	4	(1)	(1)
0.35	12	7	5	(2)	(1)
0.25	12	5	5	(2)	(1)
0.15	10	7	(4)	(3)	(1)
0.05	28	20	(12)	(5)	(2)

Table 2   System sizes obtained by explicit construction according to specified thresholds δ and radii ε for a single shift graph with $\alpha_0 = 1/2$.

$\delta \setminus \varepsilon$	0.400	0.450	0.490	0.495	0.499
0.0100	9	3	1	1	1
0.0050	9	9	3	1	1
0.0010	27	9	3	5	1
0.0005	27	9	3	3	3
0.0001	27	27	9	3	3

One advantage of the three-step construction of Theorem 5 is that the first step, the opening game, removes any bias in the graph screening function.[12] By application of the first step in the design of a reliable system, the graph screening function is moved over such that it intersects the value $1/2$ with arbitrary precision. After application of this procedure, the screening of the system is unbiased. Figure 5 illustrates how a biased agent screening function can be repaired. With a few agents (4 in the example), any level of judgemental bias can be removed – or added, if the bank prefers a more conservative bias. The example further shows how an unbiased system, G, can be steepened by applying the middle and end game.

In the example, the unbiased graph G is steepened by one composition with the selfdual graph G_*. This requires a system including a total of 20 agents.[13] This example is a forceful demonstration that the procedures provided in the present study can be used to achieve incremental improvements even for small economic systems. In bank F, such design improvements will have the likely effect of decreasing the unacceptable defaults.

More generally, table 1 shows the number of agents needed to produce an unbiased graph screening function. These numbers were obtained from application of the opening game of Theorem 5, i.e. by construction of suitable stair graphs. Table 1 shows improvements of reliability in terms of the *reduced* graph screening function $\alpha \equiv f(\mathbf{x})$, assuming that the bank employs fairly reliable agents (an assumption supported in our case study). A reliable agent has a high value of ε because its agent screening function maps alternatives onto a probability that is close to zero (alternatives with negative value) or one (alternatives with negative value). Fewer agents are required to repair a system with a high level of ε.

---

[12] Still assuming a reasonable overlap between the zone of uncertainty and level of tolerance, the graph screening will not be much distorted, only shifted. If this assumption does not hold, the two remaining steps of the construction procedure must also be applied in order to (re-)steepen the screening function.

[13] If the agents do not have fixed roles but are assigned as evaluations are needed then 12 (the maximal evaluation count) agents will suffice.

In order to increase the reliability of the system, any initial bias must first be removed by the procedure outlined in the opening game of Theorem 5. Further improvements can then be achieved by increasing the system's discriminating ability. The middle and end game of Theorem 5 provides the procedure to achieve a desired level of discriminating ability by improving an unbiased graph screening function. We now illustrate how the results of the present paper can be used to accomplish this. Table 2 shows the number of agents that is necessary to steepen a symmetric (unbiased) screening function ($\alpha_0 = 1/2$) to various desired levels of reliability.

As tables 1 and 2 illustrate, it is possible to achieve dramatic improvement in reliability through a proper design of an evaluation structure. Even if a large number of agents are needed in order to reach a high level of reliability when starting from incompetent agents, it is always possible to obtain significant incremental improvement with a handful of agents, either by diminishing the bias or by narrowing of the agent's zone of uncertainty. Overall, our example shows how the human version of Moore-Shannon's theorem can be used to design reliable economic systems. In actual practice, the application in bank F led to a number of proposed improvements. Other applications of immediate relevance include those where a system of evaluators consider a high number of comparable projects such as insurance companies, military intelligence, and medical treatment.

5.   Conclusion

The present article contributes in a number of ways to the body of knowledge on the optimal design of economic and social organizations. A new version of the main proofs and results of Moore and Shannon (1956a,b) are provided for human agents that can make independent deliberate choices. As far as we know, previous research has not considered the possibility that evaluation structures of any size can be designed to minimize the contribution of error from human agents with powers of deliberation that are much greater than those of electric components. Our main result was that perfectly reliable organizations can be built out of such human agents. We provide a unifying method to design reliable economic systems on the basis of an adapted and enhanced version of Moore & Shannon's work (Moore and Shannon 1956a,b).

Our results hold for human agents with powers to individually accept and reject alternatives on behalf of the economic system and choose a channel of transmission and a range of the possible receivers. In addition, our results are valid even in the case of non-monotone agent screening or non-monotone graph screening. Moore & Shannon's agents had no powers to individually accept and reject alternatives on behalf of the economic system and they had no choice of possible receivers. Moore & Shannon's results further assumed a monotone agent screening and a monotone graph screening. The advantage of establishing Moore & Shannon's results also for human agents is that these methods can be used for theory development and applications relating to strategic organizational design.

The methods outlined in the present article provide answers to the two design questions: How many fallible decision makers are required to build an economic system of a certain level of reliability? How should the system be designed? Our results can be used to achieve design improvements, for example, in banks or insurance companies by removing judgemental bias, increasing the unbiased judgemental ability of an evaluation structure or both. As illustrated in our case study of credit evaluation in bank F, the methods provided in the present work can be used to achieve incremental improvements even for small systems. The implications are profound in view of the significant improvements that may be achieved in high-reliability systems even with a handful of agents.

More generally, decision makers in boards, corporate headquarters and management teams, often base their selection of assets and R&D projects on joint evaluation. In the case of imperfect markets,

the design of a proper evaluation structure can remove unwarranted bias and increase the firm's judgemental ability beyond that of the individual decision maker. The implication is that the design of evaluation structures provides part of the answer to the problem of asset evaluation in incomplete markets, both in practice and theory.

The present article highlights the importance of understanding the properties of evaluation structures that are used in decision making structures in business organizations and a host of other social and economic systems. The components of the current framework can readily be used practical applications and theoretical work (see e.g. Knudsen and Levinthal (2007)) in our field. Some limitations of the present contribution deserve to be highlighted, including our focus on homogenous agents. The methods provided in the present article can readily be extended to address a number of interesting cases, including heterogeneity in ability, endogenous emergence of ability as a function of location in a decision making structure, and choice among more than two alternatives. We expect to see future work elaborating on these issues.

Acknowledgments

References

Balakrishnan, N., C.R. Rao (eds.). 2001. *Handbook of Statistics, Vol. 20. Advances in Reliability.* Elsevier, New York.

Ben-Yashar, R., S. Nitzan. 1997. The optimal decision rule for fixed size committees in dichotomous choice situations: The general result. *International Economic Review* **38**(1) 175–187.

Bogner, M.S. (ed.). 1994. *Human Error in Medicine.* Erlbaum, Mahwah, NJ.

Bolton, P., M. Dewatripont. 1994. The Firm as a Communication Network. *Quarterly Journal of Economics* **109** 809–839.

Burton, R.M., B. Obel. 1984. *Designing Efficient Organizations: Modelling and Experimentation.* North-Holland, Amsterdam and New York.

Carlsson, S., U. Grenander. 1966. Some Properties of Statistical Reliability Functions. *The Annals of Mathematical Statistics* **37**(4) 826–836.

Christensen, M., T. Knudsen. 2002. *The Architecture of Economic Organization: Toward a General Framework.* Mimeo, University of Southern Denmark, Odense.

Dahiya, S., A. Saunders, A. Srinivasan. 2003. Financial Distressand Bank Lending Relationships. *The Journal of Finance* **58**(1) 375–399.

Donaldson, L. 1995. *Contingency Theory.* Dartmouth, Brookfield USA, Singapore, Sydney.

Edmondson, A.C. 1996. Learning From Mistakes Is Easier Said than Done: Group and Organizational Influences on the Detection and Correction of Human Error. *The Journal of Applied Behavioral Science* **32**(1) 5–28.

Gehrig, T., P. Regibeau, K. Rockett. 2000. Project Evaluation and Organizational Form. *Review of Economic Design* **5** 387–407.

Goldstine, H.H. 1961. Information Theory. *Science* **133**(3462) 1395–1399.

Gradstein, M., S. Nitzan. 1988. Participation, Decision Aggregation and Internal Information Gathering in Organizational Decision Making. *Journal of Economic Behavior & Organization* **10**(4) 415–431.

Grofman, B., G. Owen, S. Feld. 1983. Thirteen Theorems in Search of the Truth. *Theory and Decision* **15** 261–278.

Hess, J.D. 1983. *The Economics of Organization.* North-Holland, Amsterdam and New York.

Ioannides, Y.M. 1987. On The Architecture of Complex Organizations. *Economics Letters* **25** 201–206.

Klevorick, A.K., M. Rothschild, C. Winship. 1984. Information Processing and Jury Decision Making. *Journal of Public economics* **23** 245–278.

Knudsen, T., Levinthal, D.A. 2007. Two Faces of Search: Alternative Generation and Alternative Evaluation. *Organization Science* **18**(1) 39–54.

Koh, W.T.H. 1992. Variable Evaluation Costs and the Design of Fallible Hierarchies and Polyarchies. *Economic Letters* **38** 313–318.

Koh, W.T.H. 1992. Human Fallibility and Sequential Decision-Making: Hierarchy versus Polyarchy. *Journal of Economic Behavior and Organization* **18** 317–345.

Koh, W.T.H. 1994. Making decisions in committees: A human fallibility approach. *Journal of Economic Behavior and Organization* **23** 195–214.

Koh, W.T.H. 1994. Fallibility and Sequential Decision Making. *Journal of Institutional and Theoretical Economics* **150**(2) 362–374.

Koh, W.T.H. 2003. *On Optimal Sequential Architectures and Management Organizations.* Paper presented at the June 2003 conference on Economic Architecture arranged by LINK and the Department of Marketing, University of Southern Denmark, Odense.

Kohn, L.T., J.M. Corrigan, M.S. Donaldson. 1999. *To Err Is Human: Building a Safer Health System.* Committee on Quality of Health Care in America, Institute of Medicine. National Academy Press, Washington, D.C.

Levinthal, D.A. 1997. Adaptation on Rugged Landscapes. *Management Science* **43**(7) 934–950.

Li, H., S. Rosen, S. Wing. 2001. Conflicts and Common Interests in Committees. *American Economic Review* **91**(5) 1478–1477.

Lomnicki, Z.A. 1973. Some Aspects of the Statistical Approach to Reliability. *Journal of the Royal Statistical Society. Series A (General)* **136**(3) 395–420.

Lynn, N., N. Singpurwalla, A. Smith. 1998. Bayesian Assessment of Network Reliability *SIAM Review* **40**(2) 202–227.

Marschak, J., R. Radner. 1972. *Economic Theory of Teams.* Yale University Press, New Haven.

Moore, E.F., C.E. Shannon. 1956. Reliable circuits using less reliable relays, part I. *Journal of the Franklin Institute* **262**(Sept.) 191–208.

Moore, E.F., C.E. Shannon. 1956. Reliable circuits using less reliable relays, part II. *Journal of the Franklin Institute* **262**(Oct.) 281–297.

Von Neumann, J. 1956. Probabilistic Logic. C.E. Shannon and J. McCarthy (eds.). *Automata Studies.* Princeton University Press, Princeton.

Nitzan, S., J. Paroush. 1985. *Collective Decision Making: An Economic Outlook.* Cambridge University Press, Cambridge.

Radner, R. 1993. The Organization of Decentralized Information Processing. *Econometrica* **61**(5) 1109–1146.

Reason, J. 1990. The Contribution of Latent Human Failures to the Breakdown of Complex Systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **327**(1241) 475–484.

Reason, J. 2000. Human Error: Models and Management, *British Medical Journal* **320** 768–770.

Rivkin, J.W., N. Siggelkow. 2003. Balancing Search and Stability: Interdependencies Among Elements of Organizational Design. *Management Science* **49** 290–311.

Sah, R., J. Stiglitz. 1985. The Theory of Economic Systems: Human Fallibility and Economic Organization. *American Economic Review Papers and Proceedings* **75** 292–297.

Sah, R., J. Stiglitz. 1986. The Architecture of Economic Systems: Hierarchies and polyarchies. *American Economic Review* **76** 716–727.

Sah, R., J. Stiglitz. 1988. Committees, Hierarchies and Polyarchies. *Economic Journal* **98** 451–470.

Sah, R. 1991. Fallibility in Human Organizations and Political Systems. *Journal of Economic Perspectives* **5** 67–88.

Shapley, L.S., B. Grofman. 1984. Optimizing Group Judgmental Accuracy in the Presence of Interdependence. *Public Choice* **43** 329–343.

Stiglitz, J. 2002. Information and the Change in the Paradigm in Economics, *American Economic Review* **92**(3) 460–501.

Thompson, J.D. 1967. *Organizations in Action*. McGraw-Hill, New York.

Visser, B. 2000. Organizational communication structure and performance. *Journal of Economic Behavior & Organization* **42** 231-252.

Van Zandt, T. 1999. Real-time decentralized information processing as a model of organizations with boundedly rational agents. *The Review of Economic Studies* **66**(228) 633–658.

Proofs of theorems follow in E-companion below.

Proofs of Theorems
The proofs of all the theorems are included here.

EC.1.   The Proofs of Extremity, Optimality and Shifting
The theorems of section 3 are proved here.

EC.1.1.   The Proof of Extremity
The theorem of extremity is remarkably general. Here the theorem is proved as it is presented, and then the requirements on the agents are loosened considerably. In homogeneous graphs the agents are all exactly the same, i.e. modelled after a prototypical agent, but as the theorem is formulated homogeneity is an unnecessarily strict requirement. The agents may differ (for example in cost or types of dispatch channels) as long as they have the same screening capabilities, they may even have a shared global memory. More importantly, any number of edge types are allowed and these edges can have dynamic weights that may depend on the project description or on local non-distributed memory which can in fact distinguish the agents. The cardinality of the possible sets of graphs spanned by this theorem is enormous, and even further extensions are discussed later.

THEOREM 1. *For any positive integer n, let $\mathcal{G}_n$ be the set of all graphs that can be constructed from agents indistinguishable with respect to project screening and under common structural and dynamic requirements, such that the maximal evaluation count is no more than n. Let $P_n \in \mathcal{G}_n$ be the polyarchy and $H_n \in \mathcal{G}_n$ be the hierarchy of n members, having graph screening functions F_{P_n} and F_{H_n}, respectively. Any graph $G \in \mathcal{G}_n$ with graph screening F_G satisfies*

$$F_{H_n}(\mathbf{x}) \leqslant F_G(\mathbf{x}) \leqslant F_{P_n}(\mathbf{x}) \tag{EC.1}$$

for all $\mathbf{x}$.

Proof of Theorem 1.   The proof runs by induction on the maximal evaluation count n. The basic step, $n = 1$, is trivial as all graphs in $\mathcal{G}_1$ has the same screening function as P_1 and H_1, both consisting only of one single agent. The induction hypothesis consists of the assumption that the theorem holds for $\mathcal{G}_k$ for all k from 1 and up to some positive n. So by assumption, all graphs in $\mathcal{G}_k$ have their screening functions bounded by those of P_k and H_k. The recursiveness of the theorem will now be shown by construction.

Consider any graph $G \in \mathcal{G}_{n+1}$ of maximal evaluation count $n + 1$, and let $m > 0$ be the number of agents in the architecture. All directed edges from any node to any other node can be collected into two effective edges without affecting the screening capabilities, a rejection edge and an acceptance edge having weights equal to the collective chance of moving between the two nodes in case of rejection and acceptance respectively. Furthermore, only graphs with one entry point from the external node $\mathbf{I}$ need to be considered since other graphs will have screening functions that are linear combinations of the ones of these single-entry graphs.

After having entered the graph at some specific node, the most general form of the graph is as illustrated in Figure EC.1. Since m agents are in the structure, the number of possible agents that can receive the project after the initial evaluation is m at most (one of them being the first agent itself) regardless of the result of the agent screening. Upon arrival at the next agent (number j with probability a_j in case of acceptance and r_j for rejection), the effective sub-graph to be seen by the project has a maximal evaluation count of n since one evaluation has already been spent. Thus, the sub-graphs $G_{a_j} \in \mathcal{G}_n$ and $G_{r_j} \in \mathcal{G}_n$ belong to the set of graphs for which the theorem holds because of the induction hypothesis. This is the cornerstone of the proof.
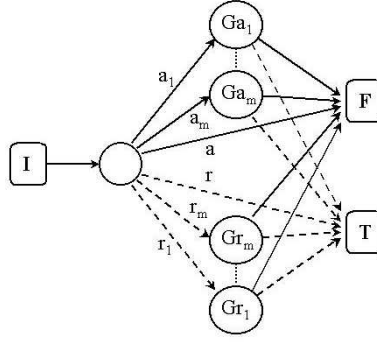
Figure EC.1 The effective form (with loops unfolded) of the general member of $\mathcal{G}_{n+1}$ immediately after entry. The solid lines are acceptance edges, the dashed lines are rejection edges and G_{a_j} (G_{r_j}) represent the sub-graph to be seen in case of acceptance (rejection) along edge j.

Besides getting skipped on to another agent, there is a possibility that the project is terminated or ultimately accepted directly as a result of the first evaluation. These probabilities are denoted r and a, respectively. As the project must leave the agent after evaluation, the weights during such a dispatching must take values in the unit interval and sum up to unity:

$$a + \sum_{j=1}^{m} a_j = 1 \qquad \wedge \qquad r + \sum_{j=1}^{m} r_j = 1 \tag{EC.2}$$

The graph screening of G can be expressed recursively in terms of the entry agent and the sub-graphs reached after the first evaluation.

$$F_G(\mathbf{x}) = f(\mathbf{x})\left(a + \sum_{j=1}^{m} a_j F_{G_{a_j}}(\mathbf{x})\right) + (1 - f(\mathbf{x})) \sum_{j=1}^{m} r_j F_{G_{r_j}}(\mathbf{x}) \tag{EC.3}$$

Similar expressions can be obtained for the polyarchy

$$F_{P_{n+1}}(\mathbf{x}) = f(\mathbf{x}) + (1 - f(\mathbf{x})) F_{P_n}(\mathbf{x}) \tag{EC.4}$$

and for the hierarchy

$$F_{H_{n+1}}(\mathbf{x}) = f(\mathbf{x}) F_{H_n}(\mathbf{x}) \tag{EC.5}$$

as well.

The recursiveness of the theorem is now established by comparing equation (EC.3) to equations (EC.4)-(EC.5). First,

$$F_{P_{n+1}}(\mathbf{x}) - F_G(\mathbf{x}) = f(\mathbf{x})\left(1 - a - \sum_{j=1}^{m} a_j F_{G_{a_j}}(\mathbf{x})\right)$$

$$+ (1 - f(\mathbf{x}))\left(F_{P_n}(\mathbf{x}) - \sum_{j=1}^{m} r_j F_{G_{r_j}}(\mathbf{x})\right) \geqslant 0 \tag{EC.6}$$

where the inequality follows from equation (EC.2) and the induction hypothesis, making both terms non-negative. Finally,

$$F_G(\mathbf{x}) - F_{H_{n+1}}(\mathbf{x}) = f(\mathbf{x})\left(a + \sum_{j=1}^{m} a_j F_{G_{a_j}}(\mathbf{x}) - F_{H_n}(\mathbf{x})\right)$$

$$+ (1 - f(\mathbf{x})) \sum_{j=1}^{m} r_j \, F_{G_{r_j}}(\mathbf{x}) \geqslant 0 \tag{EC.7}$$

is reached by similar arguments.

Invoking the principle of mathematical induction, the theorem must hold for all $n > 1$. $\square$

Dynamical dispatching of projects can be included in the model by letting the weights of the channels of communication depend on the project and the path that it has seen so far, e.g. $a_j(\mathbf{x})$. It thereby allows freelancing schemes for reusing agents and subgraphs in order to keep the total number of agents down (e.g. the consensus rule of committees), it can supply a truncation mechanism for potential infinite loops, and it allows organizations to have different modes of response triggered by certain project parameters. As mentioned, dynamical dispatching greatly extend the set of organizations that can be considered, as it adds more flexible structures with a realistic touch, usually by reducing size or cost. Despite the broader scope of such a model, the theorem of extremity still holds as it is unaffected by such arguments on the weights.

Introduction of project transformations constitute a major extension of the basic model. If agents are characterized by a (stochastic) transformation $T_{acc/rej}(\mathbf{y}, \mathbf{x})$, representing the probability that project $\mathbf{x}$ is transformed into project $\mathbf{y}$ during the evaluation process leading to acceptance/rejection, then a wide range of new models can be constructed. It then becomes possible to model imperfect channels of communication as well as examining the effects of agents actively manipulating the projects. Regardless of whether the agents transform the project before or after evaluation/dispatching, the theorem of extremity still holds. Transformations may also lift eventual degeneracies between good and bad projects of equal appearence.

Graphs build from agents that are heterogeneous in screening properties can even be treated if the graph screening performance is averaged over the distribution of different agents. This is a realistic assumption when screening performance is averaged over time in systems with a high replacement rate. If agents are regularly replaced by drawing new ones from the common agent distribution, or if each agent is actually a department of individuals whose capabilities follow the given distribution. Long time performances of organizations drawing employees from the same workforce can thus be treated, and even in these situations the theorem of extremity holds. Further extensions and a more rigorous treatment of the project selection formalism can be found in Christensen and Knudsen (2002).

THEOREM 2. *Given any threshold $0 < \delta < 1$ and a point $\alpha_0 \in\ ]0, 1[$, the number of agents n in a hierarchy such that $F_{H_n}(\alpha) \leqslant \delta$, $\forall \alpha \in [0, \alpha_0]$, is*

$$n \geqslant \frac{\log(\delta)}{\log(\alpha_0)} \tag{EC.8}$$

and the number of agents n in a polyarchy such that $F_{P_n}(\alpha) \geqslant 1 - \delta$, $\forall \alpha \in [\alpha_0, 1]$, is

$$n \geqslant \frac{\log(\delta)}{\log(1 - \alpha_0)} \tag{EC.9}$$

Proof of Theorem 2.   The result follows trivially from the graph screening functions

$$F_{H_n}(\alpha) = \alpha^n \qquad \wedge \qquad F_{P_n}(\alpha) = 1 - (1 - \alpha)^n \tag{EC.10}$$

of the n-member hierarchy H_n and polyarchy P_n. $\square$

EC.1.2. The Proof of Optimality

THEOREM 3. *For any positive integer n, let $\mathcal{D}_n$ be the set of all selfdual graphs that can be constructed from agents indistinguishable with respect to project screening and under common structural and dynamic requirements, such that the maximal evaluation count is no more than n. The slope of the screening polynomials at $\alpha = 1/2$ of the graphs in $\mathcal{D}_n$ cannot exceed that of*

$$F_{G_{n*}}(\alpha) = \sum_{i=\frac{n+1}{2}}^{n} B_i \alpha^i \quad with \quad B_i = \frac{\Gamma(n+1)}{(\Gamma(\frac{n+1}{2}))^2} \frac{(-1)^{i-\frac{n+1}{2}}}{i} \binom{\frac{n-1}{2}}{i - \frac{n+1}{2}} \quad \text{(EC.11)}$$

and at least one graph $G_{n*} \in \mathcal{D}_n$ has this reduced graph screening.

Proof of Theorem 3. This proof has two stages. First, the *polynomial view* approaches from the analytical side to find the optimal reduced screening function with respect to discriminating ability. Then the *topological view* shows by construction that a graph does exist having the found optimal screening.

The polynomial view

Graph screening functions are polynomials in the agent screening whose maximal order is the maximal evaluation count. Thus with the number of agent evaluations limited to n there are at most $n+1$ coefficients to tune. And for every (non-symmetric) constraint put on the polynomial, the selfduality constraint produces yet another. So the optimal screening polynomials should be sought within the family of odd evaluation count (assumed in the following unless otherwise stated), and at most $(n+1)/2$ conditions can be specified.

The optimality strived for here is a minimal deviation from the perfect (selfdual) screening, $\Theta(\alpha - 1/2)$, the step function. The screening is required to stay close to 0 below $\alpha = 1/2$ (and close to 1 above due to selfduality) and to change sharply around the middle of the unit interval. Clearly, the best way of achieving this is to require the screening function to be as flat as possible near the ends of the interval. In this way the screening stays closest to the extreme values for as long as possible, leaving as little as possible of the parameter space over which to perform the jump between these extremes. Moreover, this requirement ensures monotonicity which again ensures that the screening value stays in the unit interval as it must for interpretation as a probability. Whence, from a polynomial perspective the optimality requirement is that the screening and its first $(n-1)/2$ derivatives are zero at $\alpha = 0$:

$$\left. \frac{d^i F_{G_{n*}}(\alpha)}{d\alpha^i} \right|_{\alpha=0} = 0 \quad with \quad i \in \left\{ 0, 1, \ldots, \frac{n-1}{2} \right\} \quad \text{(EC.12)}$$

Here the hypothetical optimal selfdual graph with maximal evaluation count n is denoted G_{n*}.

The solution to conditions (EC.12) have a derivative proportional to α to the power $(n-1)/2$ and, by selfduality, $(1-\alpha)$ to the same power. Working out the normalization constant the entire solution (4) can be obtained from the integral of

$$F'_{G_{n*}}(\alpha) = \frac{\Gamma(n+1)}{(\Gamma(\frac{n+1}{2}))^2} \alpha^{\frac{n-1}{2}} (1-\alpha)^{\frac{n-1}{2}} \quad \text{(EC.13)}$$

with vanishing constant as $F_{G_{n*}}(0) = 0$.

The topological view

The proof is carried out in the simplest of models where each agent receives projects from a single predecessor only, and where each agent has only one successor on the acceptance and rejection side respectively. Within this model the graphs of consideration are unique. More advanced

models, for example with complex dynamical rules, can always be simplified to fit within the simple model without affecting the screening properties. This is done by duplicating, or unfolding, agents appearing on multiple paths. However, the addition of extra rules may break the uniqueness property of the graphs.

The existence (and uniqueness within the simple model) of graphs having the optimal screening is now proved by explicit construction. As they are characterized by a maximal evaluation count n and concern sets of paths of certain lengths, a few general properties of the unfolded graphs are derived first.

There is exactly one graph with fixed maximal evaluation count n that has exactly k accepts on every path to ultimate acceptance. Both existence and uniqueness can be shown by induction. Obviously $1 \leq k \leq n$ and the single agent covers the basic case of $n = 1 = k$. Assume now that the all the graphs exist up to n for all allowed k, and let (n, k) denote each of these graphs. Then $(n + 1, n + 1)$ is constructed by letting a single agent accept to (n, k), which yields the hierarchy H_{n+1}, and similarly $(n + 1, 1)$ is constructed by letting a single agent reject to $(n, 1)$, which yields the polyarchy P_{n+1}. For all intermediate k the graph $(n + 1, k)$ is constructed by letting a single agent accept to $(n, k - 1)$ and reject to (n, k). As the graphs at level n were unique, the acceptance and rejection branches are independent, and the subset of $n + 1$ graphs just constructed are all different then the uniqueness must hold as well at level $n + 1$.

Returning to the selfdual graphs of optimal screening it is seen from the zeroth derivative polynomial requirement (EC.12) that there should be at least $(n + 1)/2$ accepts on all paths to the final portfolio (again with odd e). Furthermore the selfdual variant of the same constraint dictate that there should be at most $(n + 1)/2$ accepts on all paths. Therefore there must be exactly $(n + 1)/2$ accepts on every path leading to ultimate acceptance. This is just the special graph labeled $(n, k = (n + 1)/2)$, the existence and uniqueness of which has just been proved above.   □

Although polynomials of even maximal evaluation count seem to miss a constraint compared to their odd counterparts, they are actually fully constrained as well as the selfduality requirement forces the coefficient of α^n to zero for even n. Alternatively, an additional selfduality constraint not conflicting with optimality can be put at $\alpha = 1/2$ where the generic dual constraints collapse into one, $F_{G_n}(1/2) = 1/2$. Consequently, there is nothing to gain with respect to screening capabilities by adding a single evaluation to an optimal decision structure with odd maximal evaluation count.

EC.1.3.   The Proof of Shifting

THEOREM 4. *For any $0 < \alpha_0 < 1$ and any $0 < \varepsilon < d \equiv \min(\alpha_0, 1 - \alpha_0)$, a stair graph G exists with no more than*

$$n = \lceil \log(\varepsilon/2)/\log(d) \rceil \tag{EC.14}$$

agents satisfying the relation:

$$F_G(\alpha_0 - \varepsilon) < 1/2 < F_G(\alpha_0 + \varepsilon) \tag{EC.15}$$

Proof of Theorem 4.   A sequence of stair graphs of increasing size is constructed in the following fashion. Start out with two empty dummy graphs $G_{0\downarrow}$ and $G_{0\uparrow}$ representing default strategies of rejection and acceptance respectively. Let $G_1 = A$ be the graph consisting of a single agent. These graphs obviously have screening functions, $F_{G_{0\downarrow}}(\alpha) = 0$, $F_{G_{0\uparrow}}(\alpha) = 1$, and $F_{G_1}(\alpha) = \alpha$.

For each step in the sequence, if $F_{G_n}(\alpha_0) < 1/2$ then
i) $G_{n\downarrow} = G_n$ and $G_{n\uparrow} = G_{(n-1)\uparrow}$
ii) To obtain $G_{(n+1)}$ add a new agent at rejection from the latest added agent.

else

  i) $G_{n\downarrow} = G_{(n-1)\downarrow}$ and $G_{n\uparrow} = G_n$

  ii) To obtain $G_{(n+1)}$ add a new agent at acceptance from the latest added agent.

  This sequential construction has several properties ensuring convergence, like

$$F_{G_{n\downarrow}}(\alpha_0) < 1/2 \quad , \quad F_{G_{n\downarrow}}(\alpha) \leqslant F_{G_n}(\alpha) \leqslant F_{G_{n\uparrow}}(\alpha) \quad , \quad 1/2 \leqslant F_{G_{n\uparrow}}(\alpha_0) \tag{EC.16}$$

and

$$F_{G_{n\uparrow}}(\alpha_0) - F_{G_{n\downarrow}}(\alpha_0) \leqslant d^n \quad \text{with} \quad d \equiv \max(\alpha_0, 1 - \alpha_0) < 1 \tag{EC.17}$$

While equation (3) and (4) in Moore and Shannon (1956a) does not hold for all social and economic systems and certainly not for expansion around any arbitrary agent, it does actually hold for the above sequence of graphs, which is why this specific construction is chosen. Therefore theorem 1 of the original proof, stating that

$$\frac{F'_{G_n}(\alpha)}{(1 - F_{G_n}(\alpha))F_{G_n}(\alpha)} > \frac{1}{(1 - \alpha)\alpha} \tag{EC.18}$$

(except at the endpoints of the interval and if $n = 1$ where equality holds), can be be applied to prove by contradiction that $F_{G_n}(\alpha)$ intersects $1/2$ within I whenever

$$n \geqslant \log(\varepsilon/2)/\log(d) \tag{EC.19}$$

agents have been added.   □

EC.2.   Proof of the Human Moore-Shannon Theorem and Extension
The theorems of section 4 are proved here.

EC.2.1.   The Proof of Human Moore-Shannon Theorem
  THEOREM 5. *Given any position $0 < \alpha_0 < 1$ for the shift in graph screening, any threshold $0 < \delta < 1/2$, and any radius $0 < \varepsilon < \min(\alpha_0, 1 - \alpha_0)$, then an architecture can be constructed from no more than*

$$n_{ck} \leqslant 25 \cdot \left\lceil \frac{\log(\varepsilon/2)}{\log(\min(\alpha_0, 1 - \alpha_0))} \right\rceil \cdot \left(\frac{1}{2\varepsilon}\right)^{\frac{\log(5)}{\log(11/8)}} \cdot \left(\frac{\log(3\delta)}{\log(3/4)}\right)^{\frac{\log(5)}{\log(2)}} \tag{EC.20}$$

agents whose graph screening function, F_{G_n}, fulfills condition (8).

  Proof of Theorem 5.   To ease the exposition and increase the pleasure of reading, the proof mainly elaborates on the extensions that are necessary to generalize Moore & Shannon's proof to encompass graphs that include human agents with powers to reject and accept projects on behalf of the economic system.[14]

  The present proof uses a technique similar to that of Moore and Shannon (1956a,b). The opening game accomplishes a shifting of the screening via stair graphs which removes any initial bias. The middle and end game steepens the screening using graph composition, in which the previously found graph substitutes the agents of a highly discriminating and selfdual graph.

---

[14] The original proof of Moore and Shannon (1956a,b) is readily available in the original as well as in reprint (Sloane and Wyner 1992).
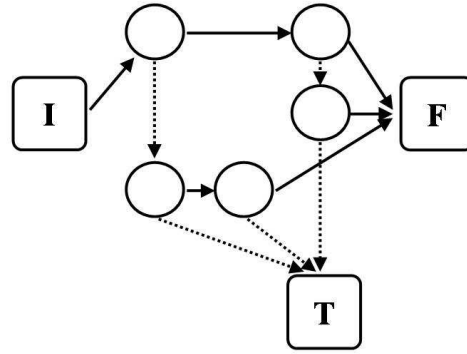
Figure EC.2   The selfdual graph G_* used in the middle and end game to steepen the graph screening function until the threshold δ is met. Full lines are acceptance edges and dashed lines are rejection edges.

The opening game.

The opening game consists of finding an architecture with n members such that the graph screening function, F_{G_n}, intersects the value $1/2$ within $I \equiv [\alpha_0 - \varepsilon, \alpha_0 + \varepsilon]$. The main difficulty lies in finding a suitable graph. Moore and Shannon (1956a,b) used ladder graphs in their proof, but these graphs cannot be used when agents have powers to reject or accept a project on behalf of the economic system. A suitable graph, including such agents, is the stair graph that is build according to the proof of Theorem 4. Thus the construction process is continued until

$$n = \lceil \log(\varepsilon/2)/\log(d) \rceil \tag{EC.21}$$

agents have been added.

The middle game.

The middle game consists in steepening the graph by recursive expansion such that $F_{G_n}(\alpha_0 - \varepsilon) < 1/4 \wedge F_{G_n}(\alpha_0 + \varepsilon) > 3/4$ is obtained. To accomplish this, begin from $G_{n\downarrow}$ and $G_{n\uparrow}$ obtained in the opening game. From $G_{n\downarrow}$ and $G_{n\uparrow}$ select the graph $G^{(0)}$ lying closest to $1/2$ at α_0 as a building block for further construction. Then select a selfdual graph to be used in recursive expansion of $G^{(0)}$. A selfdual graph in the reduced space, where the graph screening is a function of the agent screening, is defined as[15]: $F_G(\alpha) + F_G(1 - \alpha) = 1$. Recursively substituting the graph for the agents in a selfdual graph with a steep slope around $\alpha = 1/2$ will further steepen the total graph screening, first to ensure that it falls outside $[1/4, 3/4]$ on I, then to ensure that it falls outside $[\delta, 1 - \delta]$. A new sequence of graphs $\{G^{(s)}\}$ is obtained in this way. For each step, s, of this procedure of recursive expansion, the size of the graph increases. The selfdual graph G_* used here is illustrated in Figure EC.2; it is a single agent with an acceptance edge to a 2-member polyarchy and a rejection edge to a 2-member hierarchy.[16] The choice of G_* was based on the premise that it is the best (in the sense of steepening the graph screening) selfdual graph that can be obtained with a small number of agents (less than 9 agents). This optimality is guaranteed by Theorem 3 as $G_* \equiv G_{3*}$. The reduced graph screening function of the selfdual graph G_* is:

$$F_{G_*}(\alpha) = \alpha^2(3 - 2\alpha) \tag{EC.22}$$

---

[15] While the subject of selfdual is widely used, curiously little attention has been given to the study of selfdual graphs (Servatius and Christopher 1992).

[16] Although the selfdual graph used in the present article does not have as steep a slope around the center (thereby requiring more composition steps) as the 3×3 hammock networks of Moore & Shannon, it only has 5 agents (the agent count grows more slowly).

Initially $F_{G^{(0)}}(\alpha_0 - \varepsilon) < 1/2 - \varepsilon/2$ and similarly $F_{G^{(0)}}(\alpha_0 + \varepsilon) > 1/2 + \varepsilon/2$. Owing to the symmetry of the problem it suffices to consider the lower end of the interval, $\alpha = \alpha_0 - \varepsilon$. While the technique of Moore & Shannon can be applied directly, an even better bound can be obtained if the deviation γ_s from $1/2$ is observed as a function of composition count s,

$$\Delta\gamma_{s+1} \equiv 1/2 - \gamma_s - F_{G_*}(1/2 - \gamma_s) = \gamma_s(1/2 - 2\gamma_s^2) \tag{EC.23}$$

where $\gamma_0 = \varepsilon/2$ and $\gamma_{s+1} = \gamma_s + \Delta\gamma_{s+1}$. Hence the deviation from one half grows like

$$\gamma_{s+1} = \gamma_s + \gamma_s(1/2 - 2\gamma_s^2) < \frac{11}{8}\gamma_s \tag{EC.24}$$

as long as γ_s is below one quarter, which is reached no later than

$$s \geqslant -\log(2\varepsilon)/\log(11/8) \tag{EC.25}$$

compositions.

The end game.

Since $F_{G^{(s)}}(\alpha_0 - \varepsilon) < 1/4$ and $F_{G_*}(\alpha) < 3\alpha^2$ for $\alpha < 1/4$, the end game of the original proof can be applied directly, thereby showing that the desired threshold is reached in no more than

$$s \geqslant \log\left(\frac{\log(3\delta)}{\log(3/4)}\right)/\log(2) \tag{EC.26}$$

additional compositions with G_*.

The theorem is finally proven by counting the agents required by all the above steps, and a prefactor is added as some agent or composition counts may not come out even. □

EC.2.2. The Proof of the Multi-Step Theorem

THEOREM 6. *Given any threshold $0 < \delta < 1$, a series of m (odd) shift points in reduced space*

$$0 < \alpha_1 < \alpha_2 < \cdots < \alpha_m < 1 \tag{EC.27}$$

and a radius $0 < \varepsilon < \min_i(\alpha_{i+1} - \alpha_i)$, a graph can be constructed whose screening will jump from below δ to above $1 - \delta$ (and back alternatingly) within ε of the α_i's.

Proof of Theorem 6. A graph with the postulated screening function is build from the single-step functions of Theorem 5 using a sufficiently small δ' and reusing ε. As Figure 2 illustrates, the graphs shifting at the required appearences are lined up into a hierarchy, starting with α_1 closest to entry. The ones shifting from 0 to 1 must reject to the termination node, and the rest must reject to polyarchies (follows from Theorem 1) large enough to ensure almost certain acceptance as required by the threshold. In case $\alpha_1 \leqslant \varepsilon$ or $\alpha_m \geqslant 1 - \varepsilon$, the first or last single-step graph, respectively, must be replaced by a suitable polyarchy or hierarchy according to Theorem 1 and 2.

Assuming that polyarchies (and hierarchies, if needed) complete the required shift within the same δ' and ε as the generic single-step graphs, it is easy to show that the total graph will have a graph screening meeting the required threshold δ if $\delta' \leqslant \delta/m$ is used. Finally, the assumption on the polyarchies is satisfied (again according to Theorem 2) by picking $n \geqslant \lceil \log(\delta')/\log(1 - \alpha_1 + \varepsilon) \rceil$. □

References

Christensen, M., T. Knudsen. 2002. *The Architecture of Economic Organization: Toward a General Framework.* Mimeo, University of Southern Denmark, Odense.

Servatius, B., P.R. Christopher. 1992. Construction of Self-Dual Graphs. *The American Mathematical Monthly* **99**(2) 153–158.

Moore, E.F., C.E. Shannon. 1956. Reliable circuits using less reliable relays, part I. *Journal of the Franklin Institute* **262**(Sept.) 191–208.

Moore, E.F., C.E. Shannon. 1956. Reliable circuits using less reliable relays, part II. *Journal of the Franklin Institute* **262**(Oct.) 281–297.

Sloane, N.J.A., A.D. Wyner (eds.). 1992. *Claude Elwood Shannon. Collected Papers.* IEEE Information Theory Society and Wiley-Interscience, New York.