

Bureaucracy in Quest for Feasibility

Hervé Crès, Itzhak Gilboa,, Nicolas Vieille

▶ **To cite this version:**

Hervé Crès, Itzhak Gilboa,, Nicolas Vieille. Bureaucracy in Quest for Feasibility. 2012. hal-00973094

HAL Id: hal-00973094

https://sciencespo.hal.science/hal-00973094

Preprint submitted on 3 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bureaucracy in Quest for Feasibility*

Hervé Crès†, Itzhak Gilboa‡, and Nicolas Vieille§

November 2012

Abstract

The head of an organization is viewed as dealing with an optimization problem under a variety of constraints. The bureaucracy, by contrast, is viewed as dealing with the constraints alone: it has to make a multitude of low-level decisions, in such a way that no constraint is violated. However, even the feasibility problem is computationally hard. Hence bureaucracies often try to rely on past cases, in the hope of making decisions that are feasible. We study the way that past cases might affect current choices, and show that, under certain conditions, the bureaucracy will guarantee feasibility only if it mimics its behavior in a single past case.

1 Introduction

1.1 Motivation

A president of a university wishes to establish a graduate program in a new, interdisciplinary field. A program of studies is being developed and approved, and the president decides to launch it. Two years later she wishes to check

*We thank Antoine Billot, Eddie Dekel, David Schmeidler, and Yoram Weiss for conversations and references. Gilboa gratefully acknowledges ISF Grant 396/10 and ERC Grant 269754.

†Sciences-Po, Paris. herve.cres@sciences-po.fr

‡HEC, Paris, and Tel-Aviv University, tzachigilboa@gmail.com

§HEC, Paris. vieille@hec.fr

how the program is doing, and finds out that the program has not been launched yet. Inquiring why, she is told that there are staffing problems: at least until the next round of new hires, no solution was found to the assignments of instructors to the new courses as well as to the old ones. The president wonders whether it was indeed infeasible to launch the program with existing faculty, or is it the case that the administrators did not look hard enough for new solutions. There might even be a possibility that some departments were not sufficiently cooperative, perhaps because they feared losing their top students to the new program. The president has access to all the files and documents involved, including the new courses designed, the existing faculty teaching records, and so on. Yet, the complexity of the problem does not allow her to determine whether a solution could have been found.

It is often the case that decisions made within organizations are not implemented, or implemented in a different way from the directors' original intention.[1] Typically, such implementation problems will not involve explicit disobedience; rather, a decision that is made at the top level has to be translated to many minor decisions at lower levels of the hierarchy, and it is not always clear how, if at all, the top-level decision can be implemented. As a result, a decision can be stated without having any practical effects, or having results that differ from those desired.

1.2 Outline

It might be useful to think of an organization as coping with an optimization problem, where the directors set the objective function, and the bureaucracy determines the values of many decision variables subject to various feasibility constraints. According to this view, the bureaucracy consists of many lower-level decision makers, each responsible for a subset of decision variables, and

[1] Observe that we use the term "implementation" in its everyday meaning, and not in the formal sense used in the mechanism design literature.

its sole goal is to satisfy the constraints. That is, the bureaucracy consists of many decision makers who do not have a meaningful utility (or objective) function; the bureaucracy only seeks feasibility. Clearly, a feasibility problem can be viewed as a degenerate optimization problem, and, conversely, an optimization problem can be regarded as a sequence of feasibility problems. Indeed, a director of an organization can set goals for her objective function that become constraints for the bureaucracy. Yet, we find it more intuitive to think of directors as having objectives, and of the bureaucracy as seeking feasibility alone.

Given a set of constraints, how does the bureaucracy make decisions that may hopefully satisfy all constraints? In Section 2 we show that naturally-arising feasibility problems are computationally complex in a well-defined sense. Hence, it is unlikely to assume that the organization can compute an optimal solution to its optimization problem, or even that a central decision-making unit can find a feasible solution by considering the problem in the abstract. But organizations have histories, and history is often a good source of feasible solutions. For instance, if the set of constraints that the bureaucracy faces does not change over time, any solution that was chosen in the past offers a possible solution for the present. More generally, when problems faced in different periods are not necessarily identical, the bureaucracy would tend to consult past decisions in similar cases in its quest for feasibility.

However, due to practical considerations, bureaucracy has to take decisions in a decentralized way. It follows that repeating past decisions poses a non-trivial coordination problem. History offers a collection of past cases, which are relevant to the present problem to varying degrees, and different decision makers within the bureaucracy should aggregate past decisions in a way that leads to a feasible solution. In Section 3 we formulate the problem and state two possible methods of aggregation, one akin to kernel classification, and the other – to nearest-neighbor classification.

In Section 4 we show that the kernel classification method, which at-

tempts to aggregate over all similar past cases, is prone to problems akin to Condorcet's paradox. Indeed, we prove an "impossibility" theorem, along the lines of Arrow's theorem, showing that, under certain conditions, any method of aggregation that is not a (single-)nearest-neighbor approach may end up aggregating feasible solutions into an infeasible one. The case that is singled out as "the nearest" plays the role of the dictator in Arrow's theorem. However, our result does not have the same normative flavor as Arrow's impossibility theorem: first, Arrow's assumption of IIA (Independence of Irrelevant Alternatives) is replaced in our context by the assumption that the bureaucracy is constrained to make decisions in a decentralized way. This assumption is based on organizational constraints, and has no normative flavor. Second, our conclusion does not have a normatively negative meaning: while the dictatorship in Arrow's model is clearly an undesirable conclusion, using a nearest-neighbor analogy seems to be ethically neutral. Thus, our theorem bears a resemblance to social choice impossibility theorem, but its interpretation is rather different. From a mathematical viewpoint, the theorem also may be seen as a significant generalization of such impossibility results.

It is easy to observe that, if different branches of the bureaucracy have different ways of judging similarity of past cases, feasibility is not guaranteed even if all of them use a single-neighbor approach. As a result, we conclude that bureaucracies can often make decisions that violate some of the constraints. This and other implications of our model are discussed in Section 5.

1.3 Related Literature

Organizations may be thought of as monolithic, rational decision makers, maximizing expected utility under constraints. In the context of the decisions of a firm, the rational decision maker model goes back to Smith (1776), Marx (1867), and Durkheim (1893), with an emphasis on efficiency of production

in the early 20th century (Taylor, 1911, Follett, 1918, Fayol, 1919). More generally, any organization that satisfies the axioms of rational choice can be viewed as an expected utility maximizer (along the axiomatic approach of Debreu, 1959, von Neumann and Morgenstern, 1944, Savage, 1954).

However, organizations do not always seem to be sufficiently coherent to be ascribed a utility function and a subjective probability that would describe their choices via the expected utility maximization paradigm. One may espouse a different view, according to which organizations are games played by different agents, who have different utility functions, private information, and perhaps also different a priori beliefs. Decomposing organizations to sub-units was already suggested by Weber (1921, 1924), who viewed bureaucracy as a way of establishing legitimate authority, and of achieving maximal efficiency. Buchanan and Tullock (1962) viewed the state as comprising of rational agents with different goals. Niskanen (1971, 1975) analyzed bureaucracy as a production entity, and questioned its efficiency.

Our model is akin to the second strand in the literature, as it does not view the organization as a monolithic agent. However, it differs from the models mentioned above in that it does not ascribe an objective function to the bureaucracy, apart from following the path of least resistance.

Ours is by no means the only model that goes beyond the rational choice paradigm, whether applied to the organization as a whole or to components thereof. March and Simon (1958) pointed out the bounded rationality that characterizes organizational decision making. Burns and Stalker (1961) suggested that mechanistic bureaucracies are ill-adapted to deal with changing environments. Kanter (1977) argued that power inside an organization may not be easy to define, and suggested that the seemingly powerful are often powerless.[2] Bendor and Moe (1985) analyzed bureaucracies using bounded rationality models.

[2]For extensive introductions to organization theory, see Handel (2003), Hatch-Cunliffe (2006), and Scott and Davis (2007).

More generally, there are many other images that have been used to describe organizations. Morgan (2006) mentions metaphors such as machines (Taylor, 1911, Fayol, 1919, Weber, 1924), organisms (Parsons, 1951, Burns and Stalker, 1961), brains (Sandelands and Stablein, 1987, Walsh and Ungson, 1991, March, 1999), cultures (Ouchi and Wilkins, 1985), and political systems (Burns, 1961, March, 1962). For the most part, these images have not been formally modeled.

Focusing on firms, Coase (1937) pointed out the absence of an economic theory of the size of the firm. Williamson (1975, 1979, 1981) discussed the transaction costs between and within organizations, with implications for vertical integration and the boundaries of the firm. Jensen and Meckling (1976) studied modern corporation from a principal-agent point of view, highlighting the difficulties that are generated by the separation of ownership from control.

Indeed, our distinction between directors and bureaucracy brings to mind principal-agent problems. In the classical principal-agent problem (Arrow, 1963, Holmstrom, 1979), the principal does not have access to all the information available to the agent, and cannot observe the level of effort exerted by the agent. Similarly, in the problem discussed here, the principal (director) does not typically know all the bureaucratic details involved in implementing a decision, and therefore may not be able to tell whether a decision was not implemented because it could not have been implemented, or because the agent (bureaucracy) did not look hard enough for ways to implement it.

However, the director-bureaucracy problem we are interested in has several special features that distinguish it from other principal-agent problems. First, the bureaucracy need not have access to any factual information that is not observable by the principal. Rather, we assume that the principal can directly observe any file and any document that the organization possesses. Thus, the feasibility of any possible solution to the decision-implementation problem can be verified by the principal just as it can be by the bureaucracy.

The asymmetry in "information" between the bureaucracy and the principal is a result of the complexity of the implementation problem: once the principal considers a possible solution, she can verify whether it is feasible and efficacious, but a priori she cannot consider all possible solutions. By contrast, the bureaucracy can, through its collective memory, be aware of more possible solutions than can the principal.

The present paper shares much of its motivation with Gilboa and Schmeidler (2011). In particular, that paper formally models organizations as entities that make decisions without a clear utility function, and with a tendency to be consistent with past decisions. The complexity result discussed in Section 2 appeared in the first version of Gilboa and Schmeidler (2011).

The main result of the paper, presented in Section 4, is related to Arrow's impossibility theorem (Arrow, 1951) and its generalizations by Wilson (1972) and Fishburn and Rubinstein (1986), as well as to the recent literature on judgment aggregation, starting with List and Pettit (2002), and followed by, among others, Dietrich (2010), Dietrich and List (2010), Dietrich and Mongin (2010), Dokow and Holzman (2010a,b), Nehring and Puppe (2010a,b) (see List and Polak, 2010, for an introduction and a survey). Much of this literature deals with decision variables that are binary, denoting whether one alternative is preferred to another, whether a proposition is true, etc. In this context, our result extends the scope of the model to deal with non-negative integer variables that need not be restricted to $\{0,1\}$. Indeed, the proof of the theorem for the case of $\{0,1\}$ variables follows standard arguments, and the main innovation is in the extension to general variables.

2 The Complexity of Implementation of a Decision

The implementation of a decision has to take into account various resource constraints, such as monetary budgets, time constraints of employees of dif-

ferent skills, scheduling of sub-tasks, and so forth. The problem is often further complicated by the fact that the bureaucracy has several hierarchical levels, and each decision made by a given level may have to be implemented by the level below it, being translated into several minor decisions.

We show here that even if there is only one level of implementation, and the only constraints are the availability of employees, implementation may be too complex a problem to be solved by available computers and algorithmic knowledge. Specifically, we state a simple version of the implementation problem formally, and show that it belongs to the class of NP-Complete problems, for which there are no known polynomial-time algorithms.[3]

Assume that implementing a decision requires the performance of t tasks. Each task has to be performed by one of s employees. Not every employee can perform every task. Assume that the index $L_{ij} \in \{0,1\}$ denotes whether task $i \leq t$ can be performed by employee $j \leq s$. If $L_{ij} = 1$, task i would require $e_i \geq 0$ hours of employee j (assuming that all skilled employees are equally efficient in performing the task). Finally, employee $j \leq s$ has a budget of $B_j \geq 0$ hours at the organization's disposal. The implementation of the decision requires that each task i be allocated to one of the employees j such that the total hours of each employee does not exceed the available budget. An allocation can be viewed as a matrix $(a_{ij})_{i\leq t, j\leq s}$ with $a_{ij} \in \{0,1\}$. The allocation is *consistent* with $(e_i)_{i\leq t}$, $(B_j)_{j\leq s}$, and $(L_{ij})_{i\leq t, j\leq s}$ if

$$a_{ij} \leq L_{ij} \qquad \forall i,j$$

$$\sum_{j\leq s} a_{ij} = 1 \qquad \forall i$$

and

$$\sum_{i\leq t} a_{ij} e_i \leq B_j \qquad \forall j$$

That is, an allocation has to conform to the skill requirements (so that $L_{ij} = 0$ implies $a_{ij} = 0$); it has to make sure that every task is performed ($\sum_{j\leq s} a_{ij} =$

[3]See Appendix B for an informal explanation of these concepts.

1) and that no employee is asked to work more than her allowed number of hours ($\sum_{i \leq t} a_{ij} e_i \leq B_j$).

Finding such an allocation is a difficult problem:

Proposition 1 *Given expenses* $(e_i)_{i \leq t}$, $(B_j)_{j \leq s}$, *and* $(L_{ij})_{i \leq t, j \leq s}$ *finding whether there exists a consistent allocation* $(A_{ij})_{i \leq t, j \leq s}$ *is NP-Complete.*

The proof of the proposition does not make use of more than two budgets ($s = 2$) or of the matrix L (in fact, it assumes that $L_{ij} \equiv 1$).[4] Thus a problem that would appear to be much simpler would still be as hard as all the problems in the class NP. Clearly, the problem only becomes more complicated if one takes into account additional constraints, such as the timing of tasks and potential precedence constraints between them, budget constraints, and the like.

The notion of NP-Completeness captures a few aspects of the asymmetry between the director and the bureaucracy: first, the two are not asymmetric with respect to any "hard", factual information. Any concrete fact, such as whether a certain employee can perform a given task, or how much time an employee has, is known to the bureaucracy and to the director alike. Once a proposed solution to the allocation problem is suggested, both the bureaucracy and the director can verify whether it satisfies the constraints or not. However, before a concrete solution is proposed, the director will typically not be able to envisage all solutions due to their large number.

Can the bureaucracy solve the allocation problem? There are two reasons for which it might have greater computational abilities than the director. First, the bureaucracy might have a longer memory of past problems, which might resemble the current one. Second, the bureaucracy consists of many individuals, and they may be viewed as parts of a large, decentralized computing machine. While coordination is viewed as one of the chief goals of organizations (see Milgrom and Roberts, 1992), in the following section

[4] All proofs are relegated to Appendix A.

we study a model the bureaucracy as a collection of decentralized sub-units, whose coordination can only be done implicitly, by referring to past decisions.

3 Model

The bureaucracy is viewed as solving an integer programming problem, stated as

$$Ax \leq b$$

where $x = (x_1, ..., x_n)^T$ is a vector of non-negative integer-valued decision variables, A is a $m \times n$ matrix of real numbers, and b is a vector of m extended real numbers, $b_j \in \mathbb{R} \cup \{\infty\}$. Clearly, setting $b_j = \infty$ renders the constraint vacuous; allowing this possibility simplifies notation in the sequel. Evidently, constraints of the type $\leq$ can also capture constraints of the type $\geq$ and hence also equality constraints. We denote the set of possible right-hand-side (RHS) vectors by $B \equiv (\mathbb{R} \cup \{\infty\})^m$.

The restriction that all of the x's be integer-valued is not crucial and one may allow some x's to be continuous. However, complexity results as Proposition 1 do depend on integrality constraints for some variables. In particular, linear constraints can capture binary decisions, as in the task assignment problem, by using variables restricted to $\{0, 1\}$. On the other hand, variables that are generally thought of as continuous can be approximated by integer ones.[5] Thus, for simplicity, we assume that all variables are integer-valued, and all constraints are of the type $\leq$.

Recall that we use the term "solving a problem" for finding a *solution* which is no more than an assignment of values to the decision variables, satisfying all constraints. When a solution also optimizes an objective function, it is referred to as an "optimal solution". However, in this section we do not deal with objective functions.

[5] And typically are integer-valued in the final analysis.

We wish to model the bureaucracy's reliance on past decisions in its attempt to find a solution to the present problem. To this end, we need to introduce a time index, t, such that, at period t the bureaucracy faces the decision problem

$$A_t x \leq b_t$$

where x is a vector with n_t decision variables. To simplify notation, we will assume that n_t and A_t are independent of t. This involves no loss of generality because, at each period t, we can take the union of all decision variables that appeared in periods $\tau \leq t$, which is a finite set of variables. A decision variable x_i that was not actually present in period τ can be formally included in the problem with the constraint $x_i \leq 0$. Similarly, we can consider the matrix A that consists of the union of all rows of the matrices $(A_\tau)_{\tau \leq t}$ and, when a certain constraint does not appear in period τ, set $b_{j,\tau} = \infty$.

To sum, we assume that at each period t the bureaucracy faces the problem

$$Ax \leq b_t$$

with a period-independent matrix A, and period-dependent RHS vector $b_t \in B$.

Problems encountered in the past would typically have additional characteristics apart from the RHS vector. Let P be an abstract set denoting possible *circumstances* of a problem. Thus, a *problem* presents itself as a pair $(p, b) \in P \times B$ where b denotes the RHS which directly defines the problem, and p includes various relevant circumstances which may affect decisions.

A *case* is a triple $c = (p, b, x)$ where (p, b) is a problem, and $x \in \mathbb{Z}_+^n$ describes the decisions that were made in this problem. At time t, the bureaucracy faces a problem $(p_t, b_t) \in P \times B$ and needs to make a decision $x_t \in \mathbb{Z}_+^n$, so that $Ax_t \leq b_t$. When this decision is made, the organization has a *history*

$$h_t = \left((p_0, b_0, x_0), ..., (p_{t-1}, b_{t-1}, x_{t-1}), (p_t, b_t) \right).$$

Given such a history, we suppose that decision makers would think of past problems (p_i, b_i) that were similar to the current one, (p_t, b_t). Define a *similarity* function to be

$$s : (P \times B) \times (P \times B) \to \mathbb{R}_+$$

Such a similarity function can be used to identify the decision that is "most often made in these circumstances". Specifically, consider the i-th decision, namely the value of the decision variable x_i. One may go over all past cases $\tau = 0, ..., t-1$, and see which value $x_{i,\tau}$ appeared in past periods in history h_t, weigh each period by its degree of similarity to the current one, and consider the sum of these similarity values. Formally, define, for $x \in \mathbb{Z}_+$ and history h_t,

$$S(h_t, x) = \sum_{\tau=0}^{t-1} s((p_\tau, b_\tau), (p_t, b_t)) 1_{\{x_{i,\tau}=x\}}. \tag{1}$$

We say that the bureaucracy makes decisions by *aggregation of the similarity function* s if it selects x_i that is a maximizer of $S(h_t, x)$ for each i (and each t, given any h_t). This decision mode is formally equivalent to kernel classification in statistics (see Akaike, 1954, Parzen, 1962, Silverman, 1986). In classification problems, (p, b) stands for the observable characteristics, and the classifier has to guess the "correct" x based on past examples. In the present context, x is a decision made by a branch of the bureaucracy, and there is no external definition of "correctness". However, the function that is being maximized has the same structure, with the similarity function s replacing the role of the kernel function in kernel classification.

Observe that computing a maximizer of $S(h_t, x)$ requires only the knowledge of $\{x_{i,\tau}\}_\tau$ and can thus be completely decentralized: to implement such a maximal vector $(x_i)_i$, each decision x_i can be made independently of the others.

The formal analogy to classification problems brings to mind the nearest-neighbor approach (Fix and Hodges, 1951, 1952), where the classifier selects

a past case that maximizes the similarity to the present one, and makes a prediction that is the x value that was observed in that case. Correspondingly, one may select τ such that

$$s((p_\tau, b_\tau), (p_t, b_t)) \geq s((p_r, b_r), (p_t, b_t))$$

for all $r < t$, and set $x_i = x_{i,\tau}$.

Gilboa and Schmeidler (2003) provide an axiomatization of kernel classification. They consider the choice of x_i given different histories, and show that certain consistency requirements imply that there exists a similarity function s for which the choice of x_i is made to as to maximize an aggregation of the similarity function. Gilboa and Schmeidler (2011) employ this result for binary choices in the context of organization behavior. Ravid (2009) axiomatized nearest-neighbor approaches (with k neighbors, including the case of $k = 1$), and his results can similarly be re-interpreted as assumptions about the behavior of an organization.

These axiomatizations highlight the role of consistency of choice across different histories. However, in our context there is another notion of consistency that is crucial: consistency of the many minor decisions made by different branches of the bureaucracy. In particular, in the next section we ask, which decision mode guarantees that feasibility is retained.

4 Feasibility and The Prominence of a Single Case

4.1 A Condorcet-Style Problem

Consider the following example, which is a highly simplified version of the course scheduling problem discussed in the introduction. Mary can teach any one of three courses, but her teaching load is only two courses. Let $x_i \in \{0, 1\}$ denote whether she is assigned to course $i = 1, 2, 3$. Thus, the

three variables have to satisfy a single constraint,

$$x_1 + x_2 + x_3 = 2$$

(which, to fit the mold of our model, can be written as two inequalities, $x_1 + x_2 + x_3 \leq 2$ and $-x_1 - x_2 - x_3 \leq -2$). Assume that these courses belong to different programs and are run by different branches of the administration, so that each variable x_i is governed by a different office.

Next assume that we are at period $t = 3$, and that the three past problems ($\tau = 0, 1, 2$) had the same constraint. However, in each past period there have been additional constraints, for example, the presence of a visitor who was assigned to one of the courses. Be that as it may, past decisions varied across periods. Specifically, suppose that the observed past decisions are

$x_{i\tau}$	$\tau = 0$	$\tau = 1$	$\tau = 2$
$i = 1$	1	1	0
$i = 2$	1	0	1
$i = 3$	0	1	1

Assume, for simplicity, that all branches of the bureaucracy view the three past problems as equally similar to the current one. This implies that the similarity-aggregation method boils down to assigning to each variable its most common value in history. Clearly, this would lead to the assignment $x_1 = x_2 = x_3 = 1$, which violates the constraint of Mary's teaching load.

We therefore find that doing "what has been most commonly done in the past" may lead to infeasibility: even if history consists only of solutions that satisfy a certain set of constraints, choosing the most common value *for each* x_i *separately* may lead to an assignment of values which is infeasible. The analogy to Condorcet's paradox is inevitable: in this paradox, a majority vote among agents is taken, for any pair of alternatives separately. Each single voter has transitive preferences, but the majority preferences fail to be transitive. In our case, the role of the voters is played by past cases. Decision by aggregation of similarity generalizes majority vote; indeed, it

can be viewed as a (similarity-)weighted majority vote. Each single case may be feasible relative to a given constraint, while the "majority case" is not.

Similarly, the problem we face here is also akin to the "doctrinal paradox" of List and Pettit (2002). In their example, three judges have to vote on the validity of three propositions, p, q, and r. All judges agree with the "doctrine" that $((p \wedge q) \leftrightarrow r)$ that is, that r can be established if and only if both p and q hold. When the judges vote on each proposition separately, majority vote may be inconsistent with the doctrine.

The Condorcet problem suggests that "retaining the status quo" may not be as simple as it appears. Even if the bureaucracy seeks feasible solutions, independently of their optimality, following a "business as usual" approach, letting different offices do "what they usually do" may lead to inconsistent choices. A potential way out is not to attempt to summarize all of history, but simply to follow the most recent period choices, or, more generally, the most similar period choices. Assuming that the most similar period had the same set of constraints, this would guarantee feasibility. Indeed, this is the counterpart of a dictatorial solution to the aggregation problem: adopting a single agent's preferences in the social choice problem is equivalent to repeating the decisions made in a single past case. In the next sub-section we show that, under fairly general conditions, this is the most promising way to guarantee feasibility.

4.2 An Arrow-Style Theorem

Consider a given period $t \geq 2$. We will consider different histories of length t, $h_t = ((p_0, b_0, x_0), ..., (p_{t-1}, b_{t-1}, x_{t-1}), (p_t, b_t))$. Focusing on the decisions made in the past, such a history defines an $n \times t$ matrix, so that $x_{i\tau}$ is the value chosen for the variable x_i at period $\tau < t$. We assume that the decision at time t, x_{it}, is made for each i separately, as a function of past decisions made for the same variable. This separability assumption corresponds to

Arrow's (1951) IIA (Independence of Irrelevant Alternatives) assumption or the Independence assumption of List and Pettit (2002). However, in our case it is not offered as a normative axiom on social choice; rather, it is dictated by necessity: different branches of the bureaucracy have to make daily decisions in a decentralized way. Each office has access to its own decisions in the past, but not necessarily to others'. Thus, we take it as an organizational constraint that each x_i will be determined based on its own past values alone.

Assume, then, that there is a function

$$f : \mathbb{Z}_+^t \to \mathbb{Z}_+$$

such that, for each i,

$$x_{it} = f\left(x_{i0}, x_{i1}, ..., x_{i(t-1)}\right)$$

A function f will be called a *most-similar-case function* if there exists $\tau < t$ such that

$$f\left(x_0, x_1, ..., x_{t-1}\right) = x_\tau$$

for all $(x_0, x_1, ..., x_{t-1}) \in \mathbb{Z}_+^t$. It is easily seen that f is a most-similar-case function according to this definition if and only if there exists a 1-1 function $s : (P \times B) \times (P \times B) \to \mathbb{R}_+$ such that f is the nearest-neighbor decision according to the similarity function s. (Where the restriction to 1-1 functions is needed to avoid the ambiguity generated by ties.)

Clearly, even if all offices of the bureaucracy make decisions by the same most-similar-case function, feasibility is not guaranteed if the problem faced by the organization is new. In fact, the new problem may not be feasible at all. However, in the extreme case in which the same feasibility problem is repeated, and if the problem has been solved in each period in the past, one may hope that a (feasible) solution will be found again.

To capture this intuition, define a history

$$h_t = \left(\left(p_0, b_0, x_0\right), ..., \left(p_{t-1}, b_{t-1}, x_{t-1}\right), \left(p_t, b_t\right)\right)$$

to be *regular* if

(i) $b_\tau = b$ for all $\tau \leq t$;

(ii) x_τ is feasible for all $\tau < t$ (that is, $Ax_\tau \leq b$).

A function $f : \mathbb{Z}_+^t \to \mathbb{Z}_+$ is *consistent* if it generates a feasible decision for each regular history. That is, for a regular $h_t = ((p_0, b, x_0), ..., (p_{t-1}, b, x_{t-1}), (p_t, b))$, it has to be the case that

$$Ax \leq b$$

where

$$x_{it} = f\left(x_{i0}, x_{i1}, ..., x_{i(t-1)}\right).$$

To establish our result, we need to make sure that the matrix of constraints, A, allows for sufficient interaction among the decisions. Indeed, if, for example, each constraint involves only one variable, the bureaucracy's decisions may well be decentralized without fear of inconsistency. However, this would hardly be a realistic model of actual organizations. To rule out some trivial cases such as this, we assume that the matrix A satisfies the following condition.

A $(m \times n)$ matrix A contains *potentially conflicting rows* if (i) there exist two rows $i_1, i_2 \leq m$ and three columns $j_1, j_2, j_3 \leq n$ such that $a_{i_1 j} > 0$ and $a_{i_2 j} < 0$ for $j = j_1, j_2, j_3$ and (ii) there exists a row i such that $\sum_j a_{ij} > 0$.[6] In our example there was one constraint, namely,

$$x_1 + x_2 + x_3 = 2.$$

Translating this constraint to $\leq$ inequalities, it would take the form

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 2 \\ -x_1 - x_2 - x_3 &\leq -2. \end{aligned}$$

These two constraints would appear in the matrix A as

$$\begin{matrix} +1 & +1 & +1 \\ -1 & -1 & -1 \end{matrix}$$

[6] We comment on condition (ii) after the statement of the theorem.

which clearly define A as containing potentially conflicting rows. Differently put, the condition of "containing potentially conflicting rows" generalizes the example we started out with, by allowing any three positive values (not necessarily all 1) associated with any three negative values (not necessarily all -1), as long as the sum of (the entire) row with the positive values is a positive number. It turns out that this condition is sufficient to establish our result.

Theorem 2 *Assume that A contains potentially conflicting rows. Then f is consistent if and only if it is a most-similar-case function.*

This result shows that the only way that the bureaucracy can guarantee feasibility is by sticking to a single case, say, τ, out of each history h_t. As mentioned above, we can always find a similarity function $s : (P \times B) \times (P \times B) \to \mathbb{R}_+$ such that this single case in h_t be the most similar case to (p_t, b_t) in h_t. However, in practice it is likely that the similarity function that can serve as a focal point for the organization's sub-unit would be determined by recency. Thus, the theorem can be viewed as a possible explanation for a particular type of inertia: organizations may tend to do what they have done in the most recent period, as this is a simple way to coordinate on a single most-similar-case.

Comparing this result to the impossibility theorems in the social choice literature, our consistency requirement corresponds to the transitivity of preferences in Arrow (1951) or the "doctrine" in List and Pettit (2002). As mentioned above, Arrow's IIA condition, or List and Pettit's Independence assumption, are built into the definition of the function f. In this context, one might wonder why our theorem does not require a Pareto or a Unanimity assumption. Such a condition would mean, in our model, that the function f has to *retain the status quo* in the sense that $f(c, ..., c) = c$ for all $c \in \mathbb{Z}_+$. However, it turns out that this condition is implied by consistency, if the matrix A has at least one row whose sum is positive, as guaranteed by condition

(ii).

Observe that if every row of the matrix A adds up to zero, then functions that do not retain the status quo can also be consistent. Specifically, if f is a most-similar-case function and f' is defined by $f'(x_0, x_1, ..., x_{t-1}) = f(x_0, x_1, ..., x_{t-1}) + d$ for $d \in \mathbb{Z}_{++}$, then f' is also consistent.[7] This possibility is ruled out by condition (ii), which appears rather natural for constraints arising in real-life problems.

5 Discussion

Observe that in Theorem 2 the histories that are considered are not necessarily compatible with the function f discussed. Indeed, for a regular history (with a RHS vector b that is independent of the period), any function f that retains the status quo will choose the same x in each period. However, various unmodeled phenomena may yield different x's in the past. This is akin to the definition of a strategy in an extensive form game, which is defined also at nodes that are inconsistent with itself. Indeed, one would like to have the strategy defined also in case of "trembling hand" deviations, and, similarly, to have the function f defined also on histories in which past decisions ended up being different than the choice dictated by f.[8]

Real bureaucracies will face problems that are more complicated than those modeled in our Proposition 2. Typically, the RHS vectors will vary from period to period, and the new vector, b_t may not have been encountered in the past at all. If it is the case that $b_\tau \leq b_t$ for some $\tau \leq t$, one may be assured that the present problem is feasible. Indeed, introducing the inequality $b_\tau \leq b_t$ into the judgment of similarity of the problem (p_τ, b_τ) to (p_t, b_t) might mean that, when maximizing $s((p_\tau, b_\tau), (p_t, b_t))$, the bureaucracy will find

[7] It follows from the proof in the Appendix that these functions are the only consistent ones. In particular, the only consistent functions that retain the status quo are the most-similar-case ones, even when all rows of A add up to zero.

[8] Note, however, that the function f is defined for a given history length t, as opposed to a strategy that is defined for the entire game tree.

a past case τ whose solution is still feasible under present circumstances. However, it may be the case that no such past case τ exists, and then there is no guarantee that there is a practical way to find out whether a feasible solution exists.

Additional difficulties that arise in reality are that the judgment similarities of different offices of the bureaucracy need not be identical. For idiosyncratic or systematic reasons, different people would vary in their similarity judgments. In this case, a most-similar-case function might prove rather sensitive, and the overall decision vector might fail to be coherent. With this view in mind, it is possible that aggregated similarity functions might be more robust than most-similar-case ones. Be that as it may, there remains the possibility that different branches of the bureaucracy make reasonable decisions that, taken together, are incoherent.

6 Appendix A: Proofs

6.1 Proof of Proposition 1

It is straightforward that the problem is in NP. To see that it is NP-Complete, consider the following problem (which is known to be NP-Complete, see Gary and Johnson, 1979):

Problem KNAPSACK: Given integers $c_1, ..., c_n$ and b, is there a subset $K \subset \{1, ..., n\}$ such that $\sum_{i \in K} c_i = b$?

Given the input $c_1, ..., c_n, b$ for KNAPSACK, define an allocation problem with $(e_i)_{i \leq t}$, $(B_j)_{j \leq s}$, and $(L_{ij})_{i \leq t, j \leq s}$ as follows: $s = 2$, $B_1 = b$, $B_2 = \sum_{i \in K} c_i - b$. Next, $t = n$ and $e_i = c_i$. Finally, $L_{ij} = 1$ for all $i \leq n$ and $j = 1, 2$. Thus, any task can be performed by each of the two employees. If there exists $K \subset \{1, ..., n\}$ such that $\sum_{i \in K} c_i = b = B_1$, it is obvious that $\sum_{i \notin K} c_i = \sum_{i \in K} c_i - b = B_2$ and thus an allocation has been found such that no employee's hour budget is exceeded. If, on the other hand, such an allocation exists, satisfying $\sum_{i \in K} c_i \leq B_1$ and $\sum_{i \notin K} c_i \leq B_2$, both inequalities have to hold as equalities, and K solves the knapsack problem.

Clearly, the construction is done in polynomial time.□

6.2 Proof of Theorem 2

It is immediate that a most-similar-case function is consistent. We wish to prove the converse.

We let be given a $m \times n$ matrix A, a history length $t \geq 1$, and a consistent map $f : \mathbb{Z}_+^t \to \mathbb{Z}_+$.

Since A contains potentially conflicting rows, in particular it contains at least one row whose entries add up to a positive value. We show that this suffices to conclude that $f(0, \ldots, 0)$ must be equal to zero.

Lemma 1 *One has* $f(0, \ldots, 0) = 0$.

Proof. Using the notations of Section 4.2, consider a regular history h_t

in which the right-hand side b of the system of constraints is the null vector $\vec{0}_{\mathbf{R}^m}$, and in which all earlier decisions were set to zero: $x_{i\tau} = 0$ for each i and $\tau \leq t$. Since $x_i = f(x_{i0}, \ldots, x_{it}) = f(0, \ldots, 0)$ for each branch i of the bureaucracy, the decision x_i does not depend on i following the history h_t, and we denote this common value by $c \in \mathbb{Z}_+$.

Since f is consistent, the vector $\vec{c} := (c, \ldots, c)$ in $\mathbb{Z}_+^n$ solves $A\vec{c} \leq b = \vec{0}_{\mathbf{R}^m}$. Since there is at least one row of A whose entries add up to a positive number, this implies $c = 0$. ■

Let $i_1, i_2 \in \{1, \ldots, n\}$ and $j_1, j_2, j_3 \in \{1, \ldots, m\}$ be such that $a_{i_1 j} > 0$ and $a_{i_2 j} < 0$ for $j = j_1, j_2, j_3$. W.l.o.g., we take these rows and columns to be, respectively, the first two rows and the first three columns of A.

For notational simplicity, we denote the first three entries of the first two rows by a,b, and c, and $-a'$, $-b'$, and $-c'$ respectively, so that all six numbers a, b, c, a', b', and c' are positive. We will be interested in the submatrix generated by these three numbers. Specifically, for $d, d' \in \mathbf{R}^+ \cup \{+\infty\}$ we define the system $\mathcal{S}(d, d')$ in the variables $y = (y_1, y_2, y_3)$ as

$$\begin{cases} ay_1 + by_2 + cy_3 & \leq & d \\ a'y_1 + b'y_2 + c'y_3 & \geq & d'. \end{cases}$$

Thus, for $y \in \mathbb{Z}_+^3$ we say that y is a solution to $\mathcal{S}(d, d')$ if the above inequalities hold.

We will prove that f is a most-similar-case function. We will use the consistency property only through Lemma 2 below.

Lemma 2 *The map f satisfies the following property. Let $d, d' \in \mathbf{R}^+ \cup \{+\infty\}$, and, for $\tau = 1, \ldots, t$, let $y(\tau) \in \mathbb{Z}_+^3$ be a solution to the system $\mathcal{S}(d, d')$. Then the (three-dimensional) vector $(f(\mathbf{y}_1), f(\mathbf{y}_2), f(\mathbf{y}_3))$ is also a solution to $\mathcal{S}(d, d')$, where $\mathbf{y}_i$ stands for the t-dimensional vector $(y_i(\tau))_{\tau=1,\ldots,t}$.*

Proof. For each $\tau = 1, \ldots, t$, let $x(\tau)$ be the n-dimensional vector obtained by appending $n-3$ zeroes to $y(\tau)$. Set $b_1 = d$, $b_2 = -d'$, and $b_i = +\infty$

for $i > 2$, so that $x(\tau)$ solves $A\mathbf{x} \leq \mathbf{b}$ for each τ.

By Lemma 1, $f(\mathbf{x}_j) = 0$ for $j = 4, \ldots, n$. Since f is consistent, the n-dimensional vector $(f(\mathbf{y}_1), f(\mathbf{y}_2), f(\mathbf{y}_3), 0, \ldots, 0)$ solves $A\mathbf{x} \leq \mathbf{b}$ as well. ■

Given $d, d' \in \mathbf{R}^+ \cup \{+\infty\}$, and three t-dimensional vectors α, β, and γ, we will slightly abuse terminology and say that (α, β, γ) *solves* $\mathcal{S}(d, d')$ when $(\alpha(\tau), \beta(\tau), \gamma(\tau))$ solves $\mathcal{S}(d, d')$ for each τ. Lemma 2 thus says that $(f(\alpha), f(\beta), f(\gamma))$ is a solution to $\mathcal{S}(d, d')$ whenever (α, β, γ) solves $\mathcal{S}(d, d')$.

The proof proceeds in three steps. We first argue in **Step 0** that f retains the status quo. In **Step 1**, we next prove that f coincides with a most-similar-case function on the set of inputs $\{0,1\}^t$. In **Step 2**, we remove the latter restriction.

Step 0: f retains the status quo.

We here prove, as a preliminary step, that the preservation of status quo is a consequence of the consistency requirement.

Lemma 3 *One has $f(c, \ldots, c) = c$, for every $c \in \mathbb{Z}_+$.*

Proof. We apply Lemma 2 with $d = a \times c$, and $d' = a' \times c$. For each $\tau = 1, \ldots, t$, set $y(\tau) := (k, 0, 0)$, and observe that $y(\tau)$ is a solution to the system $\mathcal{S}(d, d')$. By Lemma 2 and using the fact that $f(0, \ldots, 0) = 0$, the triple $(f(c, \ldots, c), 0, 0)$ is also a solution to $\mathcal{S}(d, d')$. This implies $af(c, \ldots, c) \leq d$ and $a'f(c, \ldots, c) \geq d'$, so that $f(c, \ldots, c) = c$, as desired. ■

Step 1: The restriction of f to $\{0,1\}^t$ is a most-similar-case function

Throughout **Step 1**, we restrict inputs in $\{0,1\}^t$, and use the following piece of notation. Given a set $B \subset \{1, ..., t\}$, we denote by $\vec{1}_B \in \{0,1\}^t$ the indicator function of B. That is, $\vec{1}_B(\tau) = 1$ iff $\tau \in B$, and we similarly denote by $\vec{0}_B$ the vector defined by $\vec{0}_B(\tau) = 0$ iff $\tau \in B$. For $B = \{1, ..., t\}$, we will abbreviate $\vec{0}_B$ and $\vec{1}_B$ to $\vec{0}$ and $\vec{1}$, respectively. Note that $\vec{1}_B = \vec{0}_{\bar{B}}$ where $\bar{B}$ is the complement of B in $\{1, ..., t\}$. Any vector $\alpha \in \{0,1\}^t$ can be written

as $\alpha = \vec{1}_B$, for $B = B_\alpha := \{\, \tau \mid \alpha(\tau) = 1 \}$. We abuse notation and write $\bar{\alpha}$ to denote $\vec{0}_{B_\alpha} = \vec{1} - \alpha$.

Observe that $f(\vec{0}) = 0$ and $f(\vec{1}) = 1$ because f retains the status quo.

Lemma 4 *For every $\alpha \in \{0,1\}^t$, one has $f(\alpha) \in \{0,1\}$.*

Proof. Assume to the contrary that $f(\alpha) \geq 2$ for some $\alpha \in \{0,1\}^t$. Choose $\mathbf{y}_1 = \alpha$, $\mathbf{y}_2 = \mathbf{y}_3 = \vec{0}$. Note that $(\alpha, 0, 0)$ solves $\mathcal{S}(a, 0)$ but $(f(\alpha), f(\vec{0}), f(\vec{0})) = (f(\alpha), 0, 0)$ doesn't solve $\mathcal{S}(a, 0)$, contrary to Lemma 2. A similar contradiction is obtained if $f(\alpha) \leq -1$. ■

In the rest of Step 1 of the proof, we derive consequences of Lemma 2, with $d := \max(a+b, a+c, b+c)$ and $d' := \min(a', b', c')$. Note that the set of integer-valued solutions of $\mathcal{S}(d, d')$ in $\{0,1\}^3$ consists of all vectors in this set, apart from $(0,0,0)$ and $(1,1,1)$.

Lemma 5 *For every $\alpha \in \{0,1\}^t$, one has $f(\bar{\alpha}) = 1 - f(\alpha)$.*

Proof. Assume to the contrary that $f(\alpha) = f(\bar{\alpha}) = \delta \in \{0,1\}$ for some α. Observe that $(\alpha, \bar{\alpha}, \vec{\delta})$ solves $\mathcal{S}(d, d')$ (since $(\alpha(t), \bar{\alpha}(t), \vec{\delta}(t))$ is either $(1,0,\delta)$ or $(0,1,\delta)$). Yet, $(f(\alpha), f(\bar{\alpha}), f(\vec{\delta})) = (\delta, \delta, \delta)$ does not solve $S(d, d')$ – a contradiction. ■

Lemma 6 *f is non-decreasing on $\{0,1\}^t$ (w.r.t. the product order).*

Proof. Assume to the contrary that $f(\alpha) = 1$ and $f(\beta) = 0$ for some $\alpha \leq \beta$. Since $\alpha \leq \beta$, we know that, for every τ, if $\alpha(\tau) = 1$, then $\beta(\tau) = 1$, that is, $\bar{\beta}(\tau) = 0$. This implies that $(\alpha(\tau), \bar{\beta}(\tau), \vec{1})$ solves $\mathcal{S}(d, d')$ for each τ. Thus, $(f(\alpha), f(\bar{\beta}), f(\vec{1}))$ solves $\mathcal{S}(d, d')$ as well. Yet $f(\bar{\beta}) = 1$ by Lemma 5, so that $(f(\alpha), f(\bar{\beta}), f(\vec{1})) = (1,1,1)$ – a contradiction. ■

Lemma 7 *Let $B, C \subset \{1, ..., t\}$ be given. If $f(\vec{1}_B) = 1$ and $f(\vec{1}_C) = 1$, then $B \cap C \neq \varnothing$ and $f(\vec{1}_{B \cap C}) = 1$.*

Proof. Assume to the contrary that $f(\vec{1}_{B\cap C}) = 0$, so that $f(\vec{1}_{\bar{B}\cup\bar{C}}) = 1$. Plainly, $1_{\bar{B}\cup\bar{C}}(\tau) = 0$ as soon as $1_B(\tau) = 1_C(\tau) = 1$ and $1_{\bar{B}\cup\bar{C}}(\tau) = 1$ as soon as $1_B(\tau) = 1_C(\tau) = 0$. Therefore, $(\vec{1}_B, \vec{1}_C, \vec{1}_{\bar{B}\cup\bar{C}})$ solves $\mathcal{S}(d, d')$. Yet, $(f(\vec{1}_B), f(\vec{1}_C), f(\vec{1}_{\bar{B}\cup\bar{C}})) = (1, 1, 1)$ – a contradiction. Hence we obtain $f(\vec{1}_{B\cap C}) = 1$. This implies $B \cap C \neq \emptyset$, since $f(\vec{1}_\emptyset) = f(\vec{0}) = 0$. ■

By exchanging the roles of zeroes and ones, one gets the following version of Lemma 7.

Lemma 8 *Let $B, C \subset \{1, ..., t\}$ be given. If $f(\vec{0}_B) = 0$ and $f(\vec{0}_C) = 0$, then $B \cap C \neq \varnothing$ and $f(\vec{0}_{B\cap C}) = 0$.*

Lemma 9 *The restriction of f to $\{0, 1\}^t$ is a most-similar-case function.*

Proof. Denote by S_1 the intersection of all sets $B \subseteq \{1, ..., t\}$ such that $f(\vec{1}_B) = 1$, and by S_0 the intersection of all sets C such that $f(\vec{0}_C) = 0$. Thanks to Lemmas 7 and 8, S_0 and S_1 are non-empty. By Lemma 6, $f(\alpha) = 1$ if and only if $\alpha \geq \vec{1}_{S_1}$ and similarly, $f(\alpha) = 0$ if and only if $\alpha \leq \vec{0}_{S_0}$.

We now prove that the sets S_0 and S_1 coincide and that this common set is a singleton. Pick any element $\tau \in S_0$. Since the inequality $\vec{1}_{\{\tau\}} \leq \vec{0}_{S_0}$ does not hold, one must have $f(\vec{1}_{\{\tau\}}) = 1$ and therefore $\vec{1}_{\{\tau\}} \geq \vec{1}_{S_1}$ and it follows that $S_1 = \{\tau\}$. Since τ was an arbitrary element of S_0, for every $\tau, \tau' \in S_0$ we have $S_1 = \{\tau\} = \{\tau'\}$ and thus $\tau = \tau'$. Hence S_0 is a singleton, $S_0 = \{\tau\}$ and, as we concluded that $S_1 = \{\tau\}$, we also have $S_0 = S_1$.

Note now that, for $\delta \in \{0, 1\}$, $f(\alpha) = \delta$ as soon as $\alpha(\tau) = \delta$. That is, $f(\alpha) = \alpha(\tau)$ for every α, as desired. ■

For clarity, we will henceforth assume that the unique element of $S_0 = S_1$ is $\tau_* = 1$.

Step 2. We now remove the restriction of inputs to $\{0, 1\}^t$, and we prove that $f(\alpha) = \alpha(1)$ for every $\alpha \in \mathbb{Z}_+^t$.

We proceed by induction and prove that, for each $k \geq 1$, one has $f(\alpha) = \alpha(1)$ for every $\alpha \in [0;k]^t$. Assume thus that the latter property holds for some $k \geq 1$ (with $k = 1$ proven in Step 1), and let $\alpha \in [0;k+1]^t$ be given. The proof that $f(\alpha) = \alpha(1)$ goes by contradiction.

Assume first that $f(\alpha) > \alpha(1)$. Fix $d' = 0$ so that the second constraint in $\mathcal{S}(d,d')$ is satisfied by every $\alpha \in \mathbb{Z}_+^t$. By possibly permuting the first three columns of A, we may assume that $a \leq b, c$. Set $d = a\alpha(1) + (b+c)k$, and observe that

$$d \geq (b+c)k \geq a(k+1) \tag{2}$$

where the last inequality follows from the facts that $b, c \geq a$ and that $k \geq 1$.

Let $\beta, \gamma \in [0;k]^t$ be given by $\beta = \gamma = (k, 0, \ldots, 0)$. By the choice of d, $(\alpha(1), \beta(1), \gamma(1))$ is a solution to $\mathcal{S}(d,d')$. Moreover, for $\tau > 1$, one has

$$a\alpha(\tau) + b\beta(\tau) + c\gamma(\tau) = a\alpha(\tau) \leq a(k+1) \leq d,$$

where the first inequality follows from the fact that $\alpha \in [0;k+1]^t$ (and that $\beta(\tau) = \gamma(\tau) = 0$) and the second – from inequality (2). Thus, $(\alpha(\tau), \beta(\tau), \gamma(\tau))$ is a solution to $\mathcal{S}(d,d')$ for all $\tau \geq 1$. Hence, $(f(\alpha), f(\beta), f(\gamma))$ is a solution to $\mathcal{S}(d,d')$. Yet $f(\beta) = f(\gamma) = k$ by the induction hypothesis, and therefore, $af(\alpha) + bf(\beta) + cf(\gamma) > a\alpha(1) + (b+c)k = d$ – a contradiction.

Assume now that $f(\alpha) < \alpha(1)$. Fix $d = +\infty$ so that the first constraint in $\mathcal{S}(d,d')$ is satisfied by every $\alpha \in \mathbb{Z}_+^t$. By possibly permuting the first three columns of A, we may assume that $a' \leq b', c'$. Set $d' = a'\alpha(1)$. Let $\beta, \gamma \in [0;k]^t$ be given by $\beta = \gamma = (0, k, \ldots, k)$. By the choice of d', $(\alpha(1), \beta(1), \gamma(1))$ is a solution to $\mathcal{S}(d,d')$. As for $\tau > 1$, one has

$$a'\alpha(\tau) + b'\beta(\tau) + c'\gamma(\tau) \geq (b'+c')k \geq 2a'k \geq a'\alpha(1) = d'$$

where the last inequality follows from the fact that $\alpha(1) \leq k+1 \leq 2k$.

Hence, $(f(\alpha), f(\beta), f(\gamma))$ is a solution to $\mathcal{S}(d,d')$. Yet $f(\beta) = f(\gamma) = 0$ by the induction hypothesis, and therefore, $a'f(\alpha) + b'f(\beta) + c'f(\gamma) = a'f(\alpha) < a'\alpha(1) = d'$ – a contradiction.

7 Appendix B: Computational Complexity

A **problem** can be thought of as a set of legitimate inputs, and a correspondence from it into a set of legitimate outputs. For instance, consider the problem "Given a graph, and two nodes in it, s and t, find a minimal path from s to t". An input would be a graph and two nodes in it. These are assumed to be appropriately encoded into finite strings over a given alphabet. The corresponding encoding of a shortest path between the two nodes would be an appropriate output.

An **algorithm** is a method of solution that specifies what the solver should do at each stage. **Church's thesis** maintains that algorithms are those methods of solution that can be implemented by **Turing machines**. It is neither a theorem nor a conjecture, because the term "algorithm" has no formal definition. In fact, Church's thesis may be viewed as defining an "algorithm" to be a Turing machine. It has been proved that Turing machines are equivalent, in terms of the algorithms they can implement, to various other computational models. In particular, a PASCAL program run on a modern computer with an infinite memory is also equivalent to a Turing machine and can therefore be viewed as a definition of an "algorithm".

It is convenient to restrict attention to **YES/NO problems**. Such problems are formally defined as subsets of the legitimate inputs, interpreted as the inputs for which the answer is YES. Many problems naturally define corresponding YES/NO problems. For instance, the previous problem may be represented as "Given a graph, two nodes in it s and t, and a number k, is there a path of length k between s and t in the graph?" It is usually the case that if one can solve all such YES/NO problems, one can solve the corresponding optimization problem. For example, an algorithm that can solve the YES/NO problem above for any given k can find the minimal k for which the answer is YES (it can also do so efficiently). Moreover, such an algorithm will typically also find a path that is no longer than the specified k.

Much of the literature on computational complexity focuses on **time complexity**: how many operations will an algorithm need to perform in order to obtain the solution and halt. It is customary to count input/output operations, as well as logical and algebraic operations as taking a single unit of time each. Taking into account the amount of time these operations actually take (for instance, the number of actual operations needed to add two numbers of, say, 10 digits) typically yields qualitatively similar results.

The literature focuses on **asymptotic** analysis: how does the number of operations grow with the size of the input. It is customary to conduct **worst-case** analyses, though attention is also given to average-case performance. Obviously, the latter requires some assumptions on the distribution of inputs, whereas worst-case analysis is free from distributional assumptions. Hence the complexity of an algorithm is generally defined as the order of magnitude of the number of operations it needs to perform, in the worst case, to obtain a solution, as a function of the input size. The complexity of a problem is the minimal complexity of an algorithm that solves it. Thus, a problem is **polynomial** if there exists an algorithm that always solves it correctly within a number of operations that is bounded by a polynomial of the input size. A problem is **exponential** if all the algorithms that solve it may require a number of operations that is exponential in the size of the input, and so forth.

Polynomial problems are generally considered relatively "easy", even though they may still be hard to solve in practice, especially if the degree of the polynomial is high. By contrast, exponential problems become intractable already for inputs of moderate sizes. To prove that a problem is polynomial, one typically points to a polynomial algorithm that solves it. Proving that a YES/NO problem is exponential, however, is a very hard task, because it is generally hard to show that there does *not* exist an algorithm that solves the problem in a number of steps that is, say, $O(n^{17})$ or even $O(2^{\sqrt{n}})$.

A **non-deterministic Turing machine** is a Turing machine that allows

multiple transitions at each stage of the computation. It can be thought of as a parallel processing modern computer with an unbounded number of processors. It is assumed that these processors can work simultaneously, and, should one of them find a solution, the machine halts. Consider, for instance, the Hamiltonian path problem: given a graph, is there a path that visits each node precisely once? A straightforward algorithm for this problem would be exponential: given n nodes, one needs to check all the $n!$ permutations to see if any of them defines a path in the graph. A non-deterministic Turing machine can solve this problem in linear time. Roughly, one can imagine that $n!$ processors work on this problem in parallel, each checking a different permutation. Each processor will therefore need no more than $O(n)$ operations. In a sense, the difficulty of the Hamiltonian path problem arises from the multitude of possible solutions, and not from the inherent complexity of each of them.

The class **NP** is the class of all YES/NO problems that can be solved in **P**olynomial time by a **N**on-deterministic Turing machine. Equivalently, it can be defined as the class of YES/NO problems for which the validity of a suggested solution can be verified in polynomial time (by a regular, deterministic algorithm). The class of problems that can be solved in polynomial time (by a deterministic Turing machine) is denoted **P** and it is obviously a subset of NP. Whether P=NP is considered to be the most important open problem in computer science. While the common belief is that the answer is negative, there is no proof of this fact.

A problem A is **NP-Hard** if the following statement is true ("the conditional solution property"): if there were a polynomial algorithm for A, there would be a polynomial algorithm for any problem B in NP. There may be many ways in which such a conditional statement can be proved. For instance, one may show that using the polynomial algorithm for A a polynomial number of times would result in an algorithm for B. Alternatively, one may show a polynomial algorithm that translates an input for B to an

input for A, in such a way that the B-answer on its input is YES iff so is the A-answer of its own input. In this case we say that B is **reduced** to A.

A problem is **NP-Complete** if it is in NP, and any other problem in NP can be reduced to it. It was shown that the **SATISFIABILITY** problem (whether a Boolean expression is not identically zero) is such a problem by a direct construction. That is, there exists an algorithm that accepts as input an NP problem B and input for that problem, z, and generates in polynomial time a Boolean expression that can be satisfied iff the B-answer on z is YES. With the help of one problem that is known to be NP-Complete (**NPC**), one may show that other problems, to which the NPC problem can be reduced, are also NPC. Indeed, it has been shown that many combinatorial problems are NPC.

NPC problem are NP-Hard, but the converse is false. First, NP-Hard problems need not be in NP themselves, and they may not be YES/NO problems. Second, NPC problems are also defined by a particular way in which the conditional solution property is proved, namely, by reduction.

There are by now hundreds of problems that are known to be NPC. Had we known one polynomial algorithm for one of them, we would have a polynomial algorithm for each problem in NP. As mentioned above, it is believed that no such algorithm exists.

8 References

Akaike, H. (1954), "An Approximation to the Density Function", *Annals of the Institute of Statistical Mathematics*, **6**: 127-132.

Arrow, K. (1951), *Social Choice and Individual Values*. New York: Wiley.

Arrow, K. (1963), "Uncertainty and the Welfare Economics of Medical Care", *American Economic Review*, **53**: 91-96

Bendor, J. and T. M. Moe (1985), "An Adaptive Model of Bureaucratic Politics", *The American Political Science Review*, **79**: 755-774.

Buchanan, J. M. and G. Tullock (1962), *The Calculus of Consent: Logical Foundations of Constitutional Democracy.* Ann Arbor: The University of Michigan Press.

Burns, T. (1961), "Micropolitics: Mechanisms of Organizational Change", *Administrative Science Quarterly*, **6**: 257-281.

Burns, T. and G. M. Stalker (1961), *The Management of Innovation.* London: Tavistock.

Coase, R. H. (1937), "The Nature of the Firm", *Economica*, New Series, **IV**: 386-405.

Debreu, G. (1959). *The Theory of Value.* Cowles Foundation Monographs Series.

Dietrich, F. (2010), "The possibility of judgment aggregation on agendas with subjunctive implications", *Journal of Economic Theory*, **145**: 603-638.

Dietrich, F. and C. List (2010), "Majority Voting on Restricted Domains", *Journal of Economic Theory*, **145**: 512-543.

Dietrich, F. and Ph. Mongin (2010), "The premiss-based approach to judgment aggregation ", *Journal of Economic Theory*, **145**: 562-582.

Dokow, E. and R. Holzman (2010a), "Aggregation of Binary Evaluations", *Journal of Economic Theory*, **145**: 495-511.

Dokow, E. and R. Holzman (2010b), "Aggregation of Binary Evaluations with Abstentions", *Journal of Economic Theory*, **145**: 544-561.

Durkheim, E. (1893), *The Division of Labor in Society*. English translation, 1949, Glencoe, Il: Free Press.

Fayol, H. (1919), *General and Industrial Management*. English translation, 1949, London: Pitman.

Fix, E. and J. Hodges (1951), "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties". Technical Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, TX.

Fix, E. and J. Hodges (1952), "Discriminatory Analysis: Small Sample Performance". Technical Report 21-49-004, USAF School of Aviation Medicine, Randolph Field, TX.

Follett, M. P. (1918), *The New State: Group Organization and the Solution of Popular Government*. 1923 edition, New York: Longman, Green, and Co.

Gary, M. and D. S. Johnson (1979), *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San-Francisco, CA: W. Freeman and Co.

Gilboa, I. and D. Schmeidler (1995). "Case Based Decision Theory", *Quarterly Journal of Economics,* **110:** 605-639.

Gilboa, I. and D. Schmeidler (2003), "Inductive Inference: An Axiomatic Approach", *Econometrica*, 71: 1-26.

Gilboa, I. and D. Schmeidler (2011), "Authorization Decisions", mimeo.

Handel, M. J. (2003), *The Sociology of Organizations*. Thousand Oaks, CA: Sage Publications.

Hatch, M.-J. and A. L. Cunliffe (2006), *Organization Theory*. 2nd Edition. Oxford: Oxford University Press.

Holmstrom, B., (1979), "Moral Hazard and Observability", *Bell Journal of Economics*, **10**: 74-91

Jensen, M. C. and W. H. Meckling (1976), "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", *Journal of Financial Economics*, **3**: 305-360.

Kanter, R. M. (1977), *Men and Women of the Corporation*. New York: Basic Books.

List, C. and P. Pettit (2002), "Aggregating Sets of Judgments: An Impossibility Result," *Economics and Philosophy*, **18**: 89-110.

List, C. and B. Polak (2010), "Introduction to Judgment Aggregation", *Journal of Economic Theory*, **145**: 441-466.

March, J. G. (1962), "The Business Firm as a Political Coalition", *Journal of Politics*, **24**: 662-678.

March, J. G. (1999), *The Pursuit of Organizational Intelligence*. Malden, MA: Blackwell.

March J. G. and J. P. Olsen (1976), *Ambiguity and Choice in Organizations*. Bergen: Universitetsforlaget.

March J. G. and H. A. Simon (1958), *Organizations*. New York: John Wiley and Sons.

Marx, K. (1867), *Capital*. English Translation, 1954, Moscow: Foreign Languages Publishing House.

Morgan, G. (2006), *Images of Organization*. Thousand Oaks, CA: Sage Publications.

Milgrom, P. and J. Roberts (1992), *Economics, Organization and Management*. Prentice-Hall International.

Nehring, K. and C. Puppe (2010a), "Abstract Arrowian Aggregation ", *Journal of Economic Theory*, **145**: 467-494.

Nehring, K. and C. Puppe (2010b), "Justifiable Group Choice", *Journal of Economic Theory*, **145**: 583-602.

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Niskanen, W. (1971), *Bureaucracy and Representative Government.* New Brunswick: Aldine Transaction.

Niskanen, W. (1975), "Bureaucrats and Politicians", *Journal of Law and Economics*, **18**: 617-644.

Ouchi, W. and A. Wilkins (1985), "Organizational Culture", *Annual Review of Sociology*, **11**: 457-483.

Parsons, T. (1951), *The Social System.* New York: The Free Press.

Parzen, E. (1962), "On the Estimation of a Probability Density Function and the Mode", *Annals of Mathematical Statistics*, **33**: 1065-1076.

Ravid, D. (2010), "Axiomatic Considerations for Using k-Nearest Neighbor Classification", mimeo.

Rubinstein, A. and P. C. Fishburn (1986), "Algebraic Aggregation Theory," *Journal of Economic Theory*, **38**: 63-77.

Sandelands, L. E. and R. E. Stablein (1987), "The Concept of Organization Mind", *Research in the Sociology of Organizations*, **5**. Greenwich, CT: JAI Press.

Savage, L. J. (1954), *The Foundations of Statistics.* New York: John Wiley and Sons.

Scott, W. R. and G. F. Davis (2007), *Organizations and Organizing: Rational, Natural, and Open System Perspectives.* Upper Saddle River, NJ: Pearson Education.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis.* London and New York: Chapman and Hall.

Simon, H. A. (1947), *Administrative Behavior.* New York: McMillan.

Smith, A. (1776), *The Wealth of Nations*. 1957 edition, New York: Appleton Century Crofts.

Taylor, F. W. (1911), *The Principles of Scientific Management*. New York: Harper.

Thompson, J. D. (1967), *Organization in Action*. New York: McGraw-Hill.

Vieille, N. (2006), "A Generalized Aggregation Impossibility Result", mimeo.

Walsh, J. P. and G. R. Ungson (1991), "Organizational Memory", *Academy of Management Review*, **16**: 57-91.

Weber, M. (1921), *Economy and Society: An Outline of Interpretive Sociology*. English translation, 1968, Berkeley, CA: University of California Press.

Weber, M. (1924), *The Theory of Social and Economic Organization*. English translation, 1947, Glencoe, Il: Free Press.

Williamson, O. E. (1975), *Markets and Hierarchies*. New York: Free Press.

Williamson, O. E. (1979), "Transaction Cost Economics: The Governance of Contractual Relations", *Journal of Law and Economics*, **22**: 233-261.

Williamson, O. E. (1981), "The Economics of Organization: The Transaction Cost Approach", *American Journal of Sociology*, **87**: 548-577.

Wilson, R. (1972), "Social Choice Theory without the Pareto Principle", *Journal of Economic Theory*, **5**: 478-486.