

Revisiting the Unitary Actor Assumption:  
Toward Realistic Aggregation of Individual Preferences in Strategy  
Research\*

Felipe A. Csaszar  
Ross School of Business  
University of Michigan  
Ann Arbor, MI 48109  
fcsaszar@umich.edu

John C. Eklund  
Marshall School of Business  
University of Southern California  
Los Angeles, CA 90089  
jceklund@marshall.usc.edu

February 23, 2026

Forthcoming in *Strategy Science*

\*We thank Senior Editor Tobias Kretschmer and three anonymous reviewers for their guidance and constructive feedback throughout the review process. We are also grateful to Daniel Albert, Dylan Boynton, Javier Gimeno, Tomasz Obloj, Kramer Quist, and Brit Sharoni for their insightful comments on earlier drafts. Finally, we thank seminar participants at Wharton, ESSEC, Emory, UT Austin, Harvard Business School, the SMS Annual Conference, and the Academy of Management Meeting for their helpful suggestions. All remaining errors are our own.

## Abstract

The long-standing unitary-actor assumption in strategy research—treating firms as monolithic entities with coherent preferences—misses that organizations are coalitions of individuals with diverse and often conflicting goals. Although behavioral perspectives have challenged this assumption, the field lacks an operational method for deriving an organizational utility function from the disparate preferences of its members and the specific structures used to aggregate them. We develop a mathematical framework that (i) maps individual utility functions into choice probabilities via a random-utility model, (ii) combines those probabilities using an explicit aggregation structure (e.g., unanimity or polyarchy), and (iii) recovers an organizational utility function that rationalizes the collective behavior. This establishes organizational utility functions as operationally meaningful: they summarize and predict organizational choice, yet are generally not simple averages of members' utilities. Instead, aggregation structures systematically reshape preferences—unanimity approximates the pointwise minima of underlying utility functions, amplifying risk aversion; polyarchy approximates the pointwise maxima, promoting risk-seeking. We illustrate strategic implications in Cournot competition and principal-agent settings, showing how internal aggregation structures shift competitive and collaborative outcomes. Overall, the framework provides a parsimonious way to retrofit unitary-actor models with behaviorally grounded organizational preferences, reconciling the coalition view of the firm with rigorous and tractable strategic analysis.

**Keywords:** aggregation; utility; organizational structure; group decision-making; evaluation

# 1 Introduction

## 1.1 Questioning the Unitary Actor Assumption

Strategy research is fundamentally concerned with elucidating the factors that drive performance heterogeneity among organizations (Nag et al. 2007, Leiblein et al. 2018). A critical, yet often implicit, assumption underpinning many theories that seek to explain performance heterogeneity—such as those found in industrial organization (Tirole 1988) and search models of firms (Levinthal 1997)—is the conceptualization of organizations as *unitary actors*. For instance, industrial organization scholars typically depict firms as profit-maximizing, risk-neutral entities, whereas the literature on search typically depicts companies as singular entities navigating complex business landscapes in pursuit of high-fitness positions. However, the unitary actor assumption has been increasingly challenged by insights from the behavioral theory of the firm and the more recent literature on microfoundations of strategy (see, e.g., Cyert and March 1963, Felin et al. 2015). These bodies of work underscore that organizations are not unitary actors but are comprised of diverse actors with potentially conflicting micro-level preferences, leading to varied and complex organizational behaviors at the macro level.

The literature on aggregation has made significant strides in bridging the gap between micro and macro levels (Knudsen and Levinthal 2007, Csaszar 2013). This work has sought to infer organizational decisions by combining two elements: individuals' choice probabilities and the organization's aggregation structure. By combining these elements, researchers can predict how individual decisions coalesce into organizational outcomes. For example, if two individuals each have a 50% chance of approving a project, the probability that an organization using unanimity decision-making will approve it is only 25%. This basic example illustrates how aggregation can lead to more conservative organizational behavior compared to individual tendencies. By explicitly modeling these processes, researchers can better

understand the emergent properties of organizational behavior, offering valuable insights into the complex dynamics that shape firm-level outcomes and performance heterogeneity (we review some of this work in Section 2.4).

Despite these advances, understanding aggregation through choice probabilities is not fully satisfactory, as utility functions—not choice probabilities—are the central construct in decision-making research. Utility functions are the *lingua franca* of this field, as they help consistently specify decision-makers’ preferences, offer rich insights into characteristics such as risk aversion and preference strength, and are supported by robust empirical methods like discrete-choice analysis (McFadden 1974). Utility functions are crucial for understanding decisions at both the individual level (e.g., prospect theory research; Kahneman and Tversky 1979) and the firm level (e.g., the industrial organization and search literatures mentioned above). However, research has largely neglected how utility functions aggregate, with few exceptions assuming that group utility functions average individual utilities (Adamowicz et al. 2005, Marchand 2014); yet, it is unclear what actual aggregation process, if any, corresponds to averaging individual utilities.

Thus, a fuller treatment of aggregation calls for examining how an organization’s utility function is derived from the utility functions of its members and the aggregation structure employed. Currently, there is no robust theoretical foundation for deriving organizational utility functions, raising critical questions about the objectivity of aggregating individual utility functions, and the accuracy of such functions in reflecting an organization’s overall preferences. At stake is the validity of the unitary actor assumption upon which much of strategy research relies.

To address these challenges, we pose the following research question: How do organizational utility functions emerge from individual preferences and aggregation structures? By exploring this question, we aim to develop a theoretical framework capable of translating individual-level preferences into organizational utility functions, thereby providing a more nuanced perspective on organizational behavior and strategic decision-making.

## 1.2 Our Approach and Contribution

To answer our research question, we introduce a mathematical method to derive an organization’s utility function given its members’ utility functions and the aggregation structure they use. The key insight we use to derive the organizational utility function is to turn individuals’ utility functions into choice probabilities (i.e., screening functions); we then aggregate these probabilities using the relevant aggregation structure, and determine what organizational utility function would produce such aggregate choice probabilities (i.e., the utility function that rationalizes the organization’s behavior). Our methodology is flexible and can accommodate arbitrarily complex utility functions, any number of individual actors, and aggregation structures. For clarity, the bulk of our analyses illustrate the simplest possible aggregation structures and individuals’ utility functions. In terms of structures, we study unanimity and polyarchy. Organizations using unanimity only approve projects if all members approve; organizations using polyarchy approve projects if any single member approves (Sah and Stiglitz 1986). In terms of individuals’ utility functions, we focus on single-variable linear functions. Most of our analyses focus on two-member organizations. Once the mechanisms underlying these basic cases are understood, we illustrate how to include non-linear utility functions, utility functions that depend on multiple variables, and more than two decision makers.

By developing this theoretical framework, we make several interconnected contributions to the strategy literature. First, we demonstrate that organizational utility functions are meaningful constructs that often differ substantially from those of the underlying individuals, and that the method of aggregation fundamentally shapes these functions. This advances our understanding of organizational decision-making by providing analytical tools to predict systematic biases in organizational choices. Second, our approach enables the “retrofitting” of unitary actor theories of the firm with behaviorally plausible organizational utility functions, establishing more realistic foundations for these models. Third, we contribute to one of the

“forgotten pillars” of the behavioral theory of the firm (Gavetti et al. 2007:528)—conflict within organizations—by developing a new understanding of how conflicting preferences aggregate. Specifically, our framework resolves the apparent contradiction highlighted by Cyert and March (1963:27), who noted that “the idea of an organization goal and the conception of an organization as a coalition are implicitly contradictory.” Additionally, our approach tackles a long-standing issue in organizational theory identified by March (1962:668–671) by providing a method that is both analytically tractable and capable of representing internal organizational conflicts.

The method we outline illuminates various strategy-related phenomena by revealing how decision-making bodies like top management teams arrive at organizational choices. These teams often encompass individuals with diverse preferences (Hambrick and Mason 1984, Finkelstein et al. 2009) and who may belong to conflicting coalitions (Mithani and O’Brien 2020, Zhang and Greve 2019). Understanding this aggregation process offers insight into how senior managers’ attributes and organizational decision structures influence strategic outcomes, such as firm responses to technological change (Eggers and Park 2018), the aggressiveness of acquisition strategies (Haleblian et al. 2009), and overall organizational risk preferences (Lim and McCann 2014). Practically, our framework provides guidance for organizational design, helping leaders structure decision-making processes to better align with strategic objectives.

The paper is structured as follows. The next section reviews prior studies that have leveraged the unitary actor assumption and existing research on aggregation. We then derive our model, present illustrative results, and conclude by developing broader theoretical and practical implications of our work.

## **2 Theoretical Background**

This section situates our research within the context of previous studies. First, we illustrate how common the unitary actor assumption is in strategy research. Next, we explore

the limitations associated with this assumption. Finally, we examine existing studies on aggregation.

## 2.1 The Unitary Actor Assumption in Strategy Research

Strategy scholars frequently employ the unitary actor assumption to simplify the theoretical treatment of firms. This assumption posits that an organization can be conceptualized as a single entity with a coherent set of preferences represented by a utility function. From a decision-making perspective, this approach treats a firm, no matter how large or complex, as an individual actor striving to achieve specific goals.

The unitary actor assumption is prevalent across various theoretical frameworks in strategy research. Prominent examples include the structure-conduct-performance paradigm, the resource-based view (RBV), positioning, evolutionary economics, Schumpeterian competition, and most game-theoretic models of competition. This assumption is evident in statements that attribute decision-making to “the firm,” such as “a firm *seeks* to be unique in its industry” (Porter 1985:16), “the consuming firm *decides* to make this product” (Coase 1988:13), or “the firm *seeks* to enter an emerging market” (Tuschke et al. 2013:415) (emphases added).

Economic models of competition predominantly adopt this assumption, typically positing that firms have a singular objective: profit maximization (often coupled with risk-neutrality). The unitary actor assumption is evident in industrial organization models, spanning from Cournot (1838/1971) to contemporary research. Indeed, with the notable exception of principal–agent models, all models presented in standard introductions to industrial organization (e.g., Tirole 1988) employ the unitary actor conceptualization of firms.

The prevalence of this assumption extends beyond economics to encompass behavioral theories in strategy. For instance, the literature on organizational search often depicts the firm as a unitary actor navigating a rugged NK landscape (Levinthal 1997). In this literature, firms are portrayed as making decisions regarding which position on the landscape to explore next, with the aim of enhancing overall performance. Similar unitary actor representations

are used in other behavioral theories of the firm, including those addressing learning, routines, attention, and aspirations.

The widespread adoption of the unitary actor assumption in strategy research is convenient, as it facilitates theoretical parsimony and enables scholars to analyze firm behavior as that of a homogeneous entity pursuing well-defined objectives. However, researchers have questioned whether this simplifying assumption can accurately reflect the complexity of real organizations. The next section critically examines this assumption.

## **2.2 Challenging the Unitary Actor Assumption**

The unitary actor assumption has been challenged by three major streams of research: the behavioral theory of the firm, the microfoundations of strategy, and the economics of organizations.

In their seminal work, Cyert and March (1963) challenge the notion that firms operate as unitary actors by emphasizing the role of conflicting interests within the firm. They argue that decision-making within firms is not centralized or monolithic but rather the result of a complex process of negotiation, compromise, and conflict resolution among various internal groups, such as managers, employees, shareholders, and other stakeholders. Cyert and March's behavioral theory highlights the importance of understanding the internal dynamics and decision-making processes within firms to fully grasp firm behavior. The unitary actor assumption, which is equivalent to viewing the firm as a black box, overlooks these crucial internal processes and conflicts.

Although conflict played a central role in Cyert and March's work, subsequent research has "largely overlook[ed] the potential for conflict in decision-making" (Joseph and Gaba 2020:268). Similarly, Gavetti et al. (2007:528) describe the role of conflict within firms as a "forgotten pillar" of the Carnegie tradition.

More recently, and in a similar vein to the ideas inspiring the behavioral theory of the firm, the literature on the microfoundations of strategy has highlighted that to understand

firm performance, one must look within the firm, at the individuals and processes occurring within it (Barney and Felin 2013, Felin et al. 2015). Studies in this domain have linked characteristics of the individuals making firm decisions to firm performance. Characteristics studied include managers' cognitive capabilities (Helfat and Peteraf 2015), individuals' skills and knowledge (Loon et al. 2020), managers' heuristics (Bingham et al. 2019), acquisition experience (Meyer-Doyle et al. 2019), scientific expertise (Hess and Rothaermel 2011), and understanding of organizational routines (Aime et al. 2009).

Despite the intense focus on individual characteristics, the microfoundations literature has paid less attention to how these characteristics are aggregated to shape organization-level behaviors and outcomes (Cowen et al. 2022:6–7). Understanding this aggregation process is crucial for a comprehensive view of firm behavior, as it bridges the gap between individual actions and collective organizational outcomes (Englmaier et al. 2019).

Parallel to these behavioral and microfoundational perspectives, the economics of organizations has also formalized departures from the unitary actor assumption. Seminal work by Coase (1937) and Williamson (1975) opened the black box of the firm, viewing it as a governance structure designed to mitigate transaction costs arising from bounded rationality and opportunism. Agency theory further formalized the divergence of interests between owners and managers (Jensen and Meckling 1976, Holmstrom 1979), while property rights theory emphasized how incomplete contracts shape power dynamics between distinct parties (Grossman and Hart 1986, Hart 1995). These theories—recognized by multiple Nobel Memorial Prizes in Economic Sciences—explicitly model the firm as composed of distinct actors with disparate utility functions. However, while organizational economics rigorously analyzes conflict and the mechanisms (contracts, incentives, authority) used to align interests, it typically does not seek to derive a single aggregate utility function for the organization itself. Instead, it focuses on equilibrium outcomes of strategic interactions between members.

Despite these theoretical advances across behavioral, microfoundational, and economic traditions, strategy research still lacks an operational concept of organizational utility that

accounts for these internal complexities. While organizational economics focuses on equilibrium outcomes of strategic interactions, and behavioral scholars focus on process, neither stream typically derives a single aggregate function that can represent the firm's collective preferences. Consequently, researchers often revert to the simplistic unitary actor assumption for tractability. The rest of this section examines two different approaches that have been used to address this aggregation issue. First, scholars have sought to aggregate individual utility functions. Second, other studies have focused on aggregating the probabilities of individuals selecting a project to capture the preferences of the overall organization. Each approach has its own unique set of strengths and weaknesses that we outline below.

### **2.3 Approach 1: Aggregating Utility Functions**

This approach has been used within the marketing literature to understand the consumer behavior of groups (e.g., a family deciding on a vacation or a committee deciding on which supplies to acquire; see, e.g., Arora and Allenby 1999, Menasco and Curry 1989). This research posits that the organizational utility function resembles the average of the utility functions of the constituent individuals. Sometimes the average is replaced by a weighted average to account for situations where some of the members (e.g., the parents) have more power than other members (e.g., the children).

Although this approach does not view the organization as a black box, it shares a similar drawback with the unitary actor approach: the realism of its organizational utility function is questionable. Specifically, it *assumes* that aggregation averages individuals' utility functions. This assumption of a simple averaging function is not necessarily accurate, as the approach does not *derive* the organizational utility function from the individuals' utility functions and the actual aggregation mechanism they use. Moreover, we know that the aggregation mechanism an organization uses changes its organizational choices (Csaszar 2012), so it must inevitably also alter the organizational utility function. In contrast, simply averaging utility functions does not account for the specific aggregation mechanism used.

Some research has taken a “negative” approach toward the question of aggregating utility functions. Two objections have been raised when discussing the idea of organizational utility functions: that interpersonal utility comparisons are intractable (Hammond 1991, Binmore 2009), and that it is impossible to aggregate preferences without violating seemingly simple and reasonable conditions (i.e., Arrow’s 1951 “impossibility theorem”).

However, these difficulties need not apply to the specific problem of aggregating preferences within organizations. As (i) one can derive organizational utility functions without making interpersonal utility comparisons (e.g., rather than aggregating utilities, the first stage of our method aggregates choices) and (ii) Arrow’s (1951) theorem is about the impossibility of an *optimal* aggregation structure (where optimality is defined in terms of satisfying a basic set of reasonable conditions: unrestricted domain, independence of irrelevant alternatives, weak Pareto principle, and non-dictatorship), whereas we just seek to describe the utility function corresponding to an *actual* aggregation structure such as unanimity or polyarchy (i.e., optimality is not a requirement).

In sum, the existing approach to deriving organizational utility functions is limited in the sense that it makes the ad-hoc assumption that the organizational utility function must be an average of individuals’ utility functions. We now turn to a second approach toward aggregation, which avoids this ad-hoc assumption at the expense of not deriving utilities but choice probabilities.

## **2.4 Approach 2: Aggregating Choice Probabilities**

Because the approach of aggregating individual utility functions has significant limitations, scholars have sought to circumvent these problems by looking at how individual *choices* (rather than utilities) are aggregated. This approach stems from social choice theory’s use of voting rules (Black 1958, Condorcet 1785/1994) and was formalized by Davis (1973) through Social Decision Schemes (SDS). SDS provide a mathematical framework where individual choice probabilities across alternatives are aggregated via rules like majority, proportionality,

or unanimity to yield a group probability distribution. Subsequent work explored when aggregation improves judgments over single experts (Hogarth 1978, Clemen 1989), noting the effectiveness of simple rules like majority (Hastie and Kameda 2005), and examined how individuals use others’ opinions (Soll and Larrick 2009) or how opinions evolve socially (Friedkin and Johnsen 2011).

Distinct from, yet related to, SDS is the literature on decision-making structures initiated by Sah and Stiglitz (1986) and extended by others (e.g., Knudsen and Levinthal 2007, Christensen and Knudsen 2010, Csaszar 2012, 2013). This literature examines how hierarchies and polyarchies aggregate individual members’ probabilities of accepting opportunities based on perceived quality, shifting the focus from social choice rules to organizational design and how structures influence evaluation quality under uncertainty. Like SDS, this approach operates in probability space rather than utility space; for instance, if two members each have a 50% chance of approving a project, a unanimity structure yields only a 25% approval probability.

Instead of using utility functions, this approach relies on the concept of a “screening function,” a function that maps the outcome of a project to the probability of that project being accepted. The simplest screening function would be one that approves all projects that have a positive outcome and rejects all projects with negative outcomes (i.e., this is a step function whose probability of approval is  $s(x) = \mathbb{1}[x > 0]$ ); however, because screening by fallible individuals is imperfect, the probability of approval may only increase slowly as outcomes shift from negative to positive. Figure 1 illustrates four different screening functions.

Panel (a) shows a typical screening function, one where the probability of accepting a project increases smoothly as the outcome ( $x$ ) of a project improves. The  $x$ -axis of a screening function is the same as the  $x$ -axis in a utility function: the outcome being assessed (e.g., the quality of a project). Note that for the screening function in panel (a), a project with a negative outcome (e.g.,  $x = -1$ ) may be accepted and one with a positive outcome may be

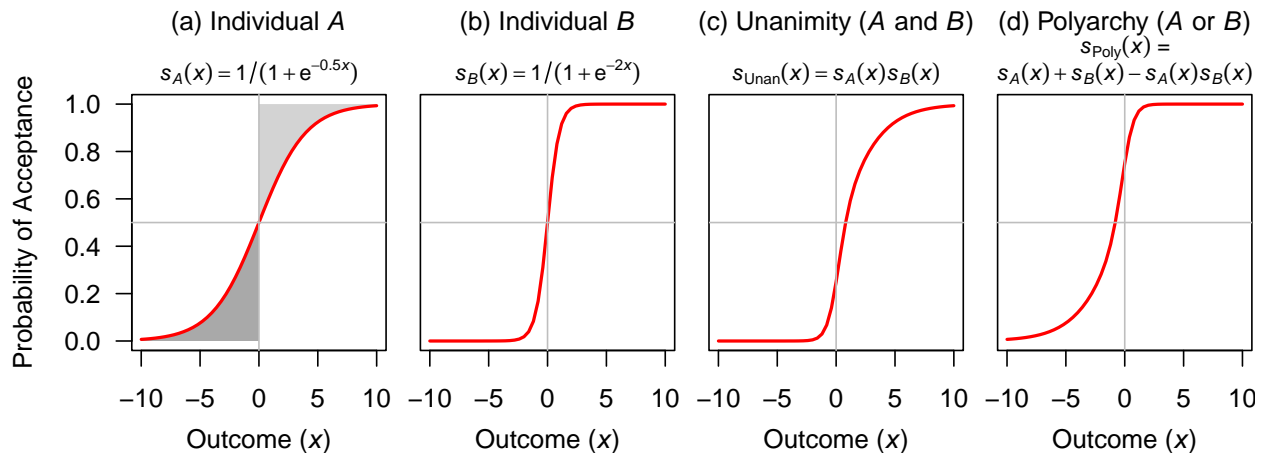


Figure 1: Examples of screening functions.

rejected. As previously mentioned, this happens because individuals are fallible: they do not see the actual outcome of a project but a noisy perception of it. The area to the right of  $x = 0$  and above the screening function (shaded in light gray in panel (a)) determines the probability of making an omission error (i.e., rejecting a good project). Similarly, the area to the left of  $x = 0$  and below the screening function (shaded in dark gray) determines the probability of making a commission error; that is, accepting a bad project.

Panel (b) illustrates a much sharper screening function, one in which the probability of accepting projects increases steeply around  $x = 0$ . The individual described in panel (b) is more accurate than the individual described in panel (a), as she is less likely to make either omission or commission errors.

Panel (c) illustrates a screening function that is shifted farther to the right than those previously described. This shift to the right means that a project needs to have a substantially better outcome to have a good chance of being accepted. Conversely, the screening function in panel (d) is shifted to the left and, hence, even projects with negative outcomes have a high chance of being approved.

One can aggregate individual-level screening functions into an organization-level screening function by using basic probability theory. For instance, an organization with two members ( $A$  and  $B$ ) that approves when  $A$  AND  $B$  approve has a screening function corresponding to

the multiplication of the members’ screening functions. So, if  $A$  and  $B$  have the screening function depicted in panels (a) and (b) respectively ( $s_A(x) = \frac{1}{1+e^{-0.5x}}$  and  $s_B(x) = \frac{1}{1+e^{-2x}}$ ), then unanimity’s screening function would be  $s_{\text{Unan}}(x) = s_A(x)s_B(x) = \frac{1}{(1+e^{-0.5x})(1+e^{-2x})}$ . Similarly, an organization that approves when  $A$  OR  $B$  approve has a screening function  $s_{\text{Poly}}(x) = s_A(x) + s_B(x) - s_A(x)s_B(x) = 1 - \frac{1}{(1+e^{0.5x})(1+e^{2x})}$ .

The customary name for an organization that approves when at least one of its members approves is *polyarchy* and the customary name for an organization that only approves when everybody approves is *unanimity* (or “hierarchy”; Sah and Stiglitz 1986, 1988, Csaszar 2012). Because unanimity requires everybody to agree, unanimity is a more conservative rule than polyarchy. Graphically, this means that the screening function of unanimity is shifted to the right vis-à-vis polyarchy (i.e., unanimity will require a higher  $x$  than polyarchy to approve projects). In fact, the curves in panels (c) and (d) are respectively the screening functions of unanimity and polyarchy derived above. Sah and Stiglitz (1988) extend these formulae to  $N$  individuals as well as to cases where the number of approvals required can be any number between 1 and  $N$ . Csaszar (2013) analyzes an even larger set of structures and Csaszar (2012) tests the predictions of Sah and Stiglitz (1986) using data from mutual funds.

The main benefit of using screening functions is that it is possible to derive an organization’s behavior from an understanding of its members and the aggregation structure used. This allows for determining useful metrics such as the probability that an organization will make omission and commission errors.

An important limitation of the screening function approach is that it does not say anything about the organization-level utility function—a key drawback given that utility functions are central to decision-making research. They offer critical insights into decision-makers’ characteristics (e.g., risk preferences) and are essential for accurately reflecting an organization’s overall preferences, which is crucial for validating the unitary actor assumption in strategy research.

Another limitation is that unlike utility functions, which have a clear connection with

individual preferences, it is unclear how screening functions represent preferences. This makes it difficult to use screening functions to examine issues related to conflict and incentives within organizations. In fact, the research on screening functions has generally avoided problems related to incentives and conflict in organizations (Joseph and Gaba 2020:268) and instead adhered to the “team-theoretic assumption” (Marschak and Radner 1972).

In sum, despite several potential benefits, prior research has not provided a way to derive an organization’s utility function from the utility functions of its members and the aggregation structure they use. The next section introduces a way of doing so.

## 3 Method

### 3.1 Overview of the Method

The key insight underlying our method is that one can avoid the difficulties of aggregating utilities by turning individual utility functions into individual screening functions, then aggregating these screening functions, and finally turning the aggregate screening function into an aggregate utility function. In shorthand, our approach moves from utility-space to probability-space, performs the aggregation there, and then comes back to utility-space.

More specifically, our approach has three steps: (1) convert the utility functions of the members of the organization into screening functions; (2) aggregate the screening functions using the method described in the previous subsection; and (3) convert the aggregate screening function into an aggregate utility function, which we call the *organizational utility function*. This last step involves looking at the organization’s behavior and asking “if this was the behavior of a person, what would be the utility function that would explain this behavior?” These three steps are illustrated as the arrows in Figure 2, which presents an overview of the main elements (for now, disregard the equations in the figure, which are explained later).

Following the literature on discrete choice (McFadden 1974), our model assumes that utility (both individual and organizational) includes a random term that captures decision-

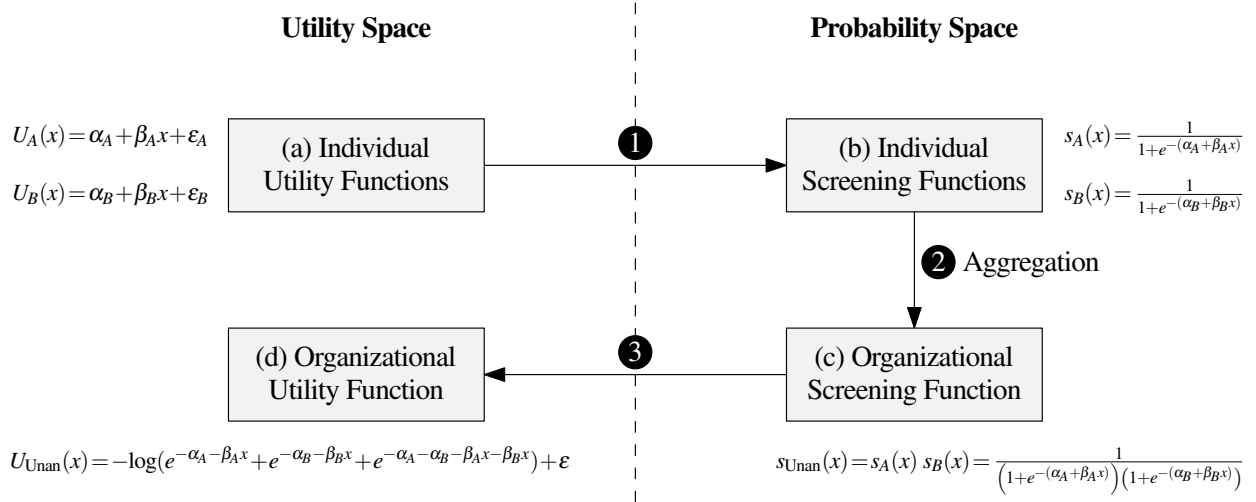


Figure 2: Summary of the method used to go from individuals’ utility functions to an organizational one.

making inconsistencies due to perceptual and cognitive limitations. For instance, the utility of being paid  $\$x$  could be  $U(x) = x^{0.5} + \varepsilon$ , where  $\varepsilon$  is a random variable.<sup>1</sup> The customary assumption regarding this random distribution is to use a Logistic(0, 1), similar in shape to the Normal distribution but analytically tractable. Note that random utility models typically keep this distribution fixed. This assumption does not make random utility models less general, as scaling the other terms of the utility function has the same effect as scaling the random term.

We say that our model is derived from first principles because it rests on two basic assumptions: (i) that individuals can be modeled using a random-utility framework, and (ii) that organizational aggregation can be described as a rule that combines members’ choice probabilities into an organizational choice probability. As shown in the previous section, both assumptions are well-established in the decision-making literature. Because our derivation depends only on these general assumptions—rather than on specific functional forms—our

<sup>1</sup>The random utility assumption originated in early psychophysics research (Fechner 1860, Thurstone 1927; see also Chapter 2 in Luce 1959) and now forms the foundation of the empirical research on decision-making (for a survey, see Hess and Daly 2014; for a recent popular treatment, see Kahneman et al. 2021). Manski (1977) identifies two distinct interpretations of the random error: psychological (representing inherent noise in human cognitive processes) and econometric (capturing researchers’ incomplete information about alternatives or decision makers). Our paper specifically addresses noisy evaluation, making the psychological interpretation most relevant to our analysis.

framework can accommodate any form of individual utility function and any aggregation structure.

Mathematically, this generality can be expressed as follows: let  $F$  denote the function that converts utilities into probabilities (arrow 1 in Figure 2),  $F^{-1}$  its inverse (arrow 3), and  $G$  the function that aggregates probabilities according to the organization’s structure (arrow 2). Then the utility function of an organization with any number of members and any aggregation structure is simply  $F^{-1}(G(F(u_1(x)), \dots, F(u_N(x))))$ . This formulation extends our method to any situation where these functions can be defined.

For clarity of exposition, we illustrate our approach using simple utility functions and aggregation structures. Specifically, we analyze single-variable linear utility functions with positive slopes, representing situations where “more is better” (i.e.,  $U(x) = \alpha + \beta x + \varepsilon$ , with  $\beta > 0$ ), and two two-member aggregation structures: unanimity and polyarchy. These simplifying choices yield tractable results and clear insights into the determinants of organizational utility functions; they can be interpreted as representing small groups like top management teams.

Later in the paper we relax these assumptions in several directions: Section 4.2 considers larger groups ( $N > 2$ ), Section 4.3 allows members with opposing preferences, Section 4.4 embeds the resulting organizational utility functions in competitive (Cournot) and collaborative (principal–agent) settings, and Appendix B explores more complex individual utility functions.

We now derive the utility functions for unanimity and polyarchy. Because much of the setup is common to both, we present most technical details in the first derivation.

## 3.2 Unanimity

We assume an organization with two members,  $A$  and  $B$ , whose utility functions are:

$$U_A(x) = \alpha_A + \beta_A x + \varepsilon_A$$

$$U_B(x) = \alpha_B + \beta_B x + \varepsilon_B$$

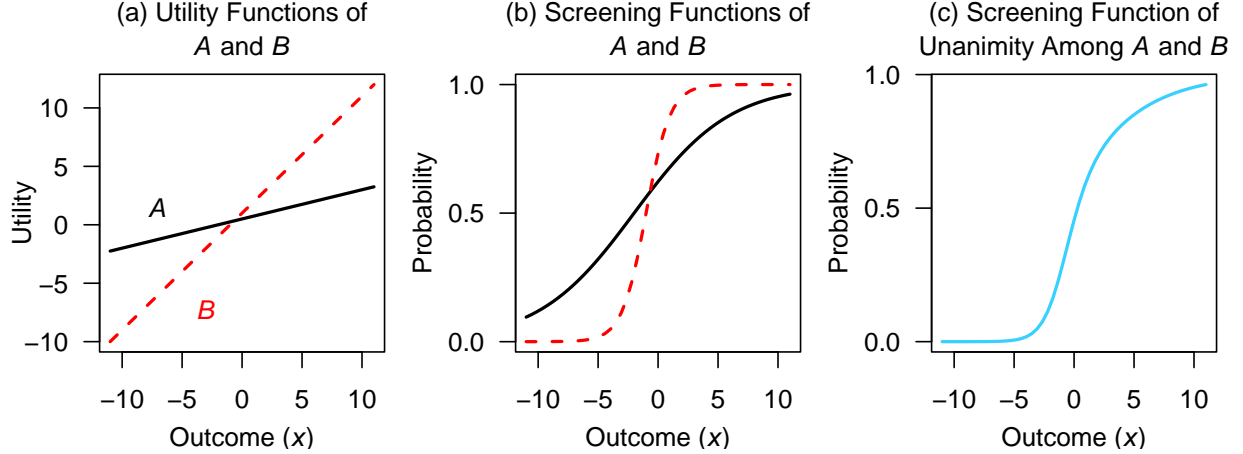


Figure 3: Illustration of the initial steps of the method: going from (a) individual utility functions to (b) individual screening functions to (c) an organizational screening function.

That is, the  $\beta$ 's control how sensitive each individual is to the value  $x$  of the project being evaluated, and the  $\alpha$ 's control their baseline level of satisfaction (e.g., an individual with a high  $\alpha$  is akin to an optimist, someone who tends to like most projects; and vice versa for low  $\alpha$ ). The  $\varepsilon$ 's are random draws from a logistic distribution with mean 0 and scale parameter 1. We assume members' evaluation errors are independent; relaxing this is straightforward but changes the aggregation algebra.

For the sake of clarity, we illustrate the model with a numerical example. We start this example with panel (a) in Figure 3, which plots the nonrandom part of the utility functions of two individuals, one with utility  $U_A(x) = 0.5 + 0.25x + \varepsilon_A$  and the other with  $U_B(x) = 1 + x + \varepsilon_B$ . In this example, individual  $B$  is both more optimistic and more sensitive than individual  $A$ .

The screening functions of these individuals—the probabilities that each recommends approval—are the probabilities that, given a project with outcome  $x$ , each individual's utility is positive. That is,

$$s_A(x) = \mathbb{P}[U_A(x) > 0] = \int_{-\infty}^{\infty} f(\varepsilon_A) \mathbb{1}[\alpha_A + \beta_A x + \varepsilon_A > 0] d\varepsilon_A = F(\alpha_A + \beta_A x) = \frac{1}{1 + e^{-(\alpha_A + \beta_A x)}}$$

$$s_B(x) = \mathbb{P}[U_B(x) > 0] = \int_{-\infty}^{\infty} f(\varepsilon_B) \mathbb{1}[\alpha_B + \beta_B x + \varepsilon_B > 0] d\varepsilon_B = F(\alpha_B + \beta_B x) = \frac{1}{1 + e^{-(\alpha_B + \beta_B x)}},$$

where  $f(\cdot)$  and  $F(\cdot)$  are, respectively, the probability density and cumulative distribution

of the logistic distribution. Thus, the screening functions of the individuals in the running example are  $s_A(x) = \frac{1}{1+e^{-(0.5+0.25x)}}$  and  $s_B(x) = \frac{1}{1+e^{-(1+x)}}$ . Panel (b) in Figure 3 plots the screening functions of these individuals.

Now that we have expressions for the screening functions of both individuals, we can compute the screening function of the organization. That is, the probability that the organization would approve a project with a given outcome  $x$ . Because unanimity approves a project when both individuals recommend approving it, the screening function of the organization is

$$s_{\text{Unan}}(x) = s_A(x) s_B(x) = \frac{1}{(1 + e^{-(\alpha_A + \beta_A x)})(1 + e^{-(\alpha_B + \beta_B x)})}. \quad (1)$$

Panel (c) in Figure 3 adds this probability to the ongoing example.

We are now in the position to ask what is the utility function that would generate the screening function of unanimity. To make this utility function comparable to those of individuals, we assume it also includes a logistic error term:  $U_{\text{Unan}}(x) = u_{\text{Unan}}(x) + \varepsilon$ , where  $\varepsilon \sim \text{Logistic}(0, 1)$ .<sup>2</sup> We know that someone with such a utility function would approve projects with probability

$$s(x) = \mathbb{P}[U_{\text{Unan}}(x) > 0] = \frac{1}{1 + e^{-u_{\text{Unan}}(x)}}. \quad (2)$$

---

<sup>2</sup>The error term in the organizational utility function can be interpreted as a form of “cognitive” error at the organizational level—that is, inconsistency or noise in the organization’s utility function and choices. This error term fully captures the inconsistency in the organization’s decision-making. Under the  $\text{Logistic}(0, 1)$  normalization for the random-utility errors (which fixes the scale), the mapping from acceptance probabilities to deterministic utility is unique:  $u(x) = \text{logit}(s(x)) = \log(s(x)/(1 - s(x)))$ . Consequently, arrow 3 in Figure 2 is reversible in the same way that arrow 1 is reversible for each individual.

Now we solve for the  $u_{\text{Unan}}(x)$  that makes equations (1) and (2) equal. That is,

$$\begin{aligned}
s(x) &= s_{\text{Unan}}(x) \\
\frac{1}{1 + e^{-u_{\text{Unan}}(x)}} &= \frac{1}{(1 + e^{-(\alpha_A + \beta_A x)})(1 + e^{-(\alpha_B + \beta_B x)})} \\
\frac{1}{1 + e^{-u_{\text{Unan}}(x)}} &= \frac{1}{1 + e^{-\alpha_A - \beta_A x} + e^{-\alpha_B - \beta_B x} + e^{-\alpha_A - \alpha_B - \beta_A x - \beta_B x}} \\
e^{-u_{\text{Unan}}(x)} &= e^{-\alpha_A - \beta_A x} + e^{-\alpha_B - \beta_B x} + e^{-\alpha_A - \alpha_B - \beta_A x - \beta_B x} \\
-u_{\text{Unan}}(x) &= \log(e^{-\alpha_A - \beta_A x} + e^{-\alpha_B - \beta_B x} + e^{-\alpha_A - \alpha_B - \beta_A x - \beta_B x}).
\end{aligned}$$

Thus, the utility function corresponding to unanimity is

$$u_{\text{Unan}}(x) = -\log(e^{-\alpha_A - \beta_A x} + e^{-\alpha_B - \beta_B x} + e^{-\alpha_A - \alpha_B - \beta_A x - \beta_B x}). \quad (3)$$

It is interesting to note that this is a complicated function, which does not look anything like the linear utility functions of the underlying individuals.

A challenge posed by this complicated functional form is that it is hard to interpret: it is not clear what this function does to the utility functions of  $A$  and  $B$ . However, this functional form—known as “LogSumExp” (a logarithm of a sum of exponentials)—has appeared in various literatures and is well understood (see, e.g., Boyd and Vandenberghe 2004:72, Lugosch 2020). The key insight is that this function closely approximates the minimum of its arguments, essentially tracing along whichever individual utility function is lowest at each point. For a detailed explanation of this approximation and its bounds, see Appendix A.

Panel (a) in Figure 4 illustrates this property, showing the utility function of unanimity (marked as a thick line) along with  $u_A(x)$ ,  $u_B(x)$ , and  $u_A(x) + u_B(x)$  for individuals with utility functions  $u_A(x) = 5 + x$  and  $u_B(x) = -5 + 3x$ . Notice how in each of the three segments of panel (a), separated by vertical lines, the organizational utility function traces the minima of the individuals’ utility functions.<sup>3</sup>

---

<sup>3</sup>All figures plotting an organizational utility function use the actual utility function  $u(x)$ , not its approximation. Given the tight approximation bounds (see Appendix A) and the plotted range, the actual

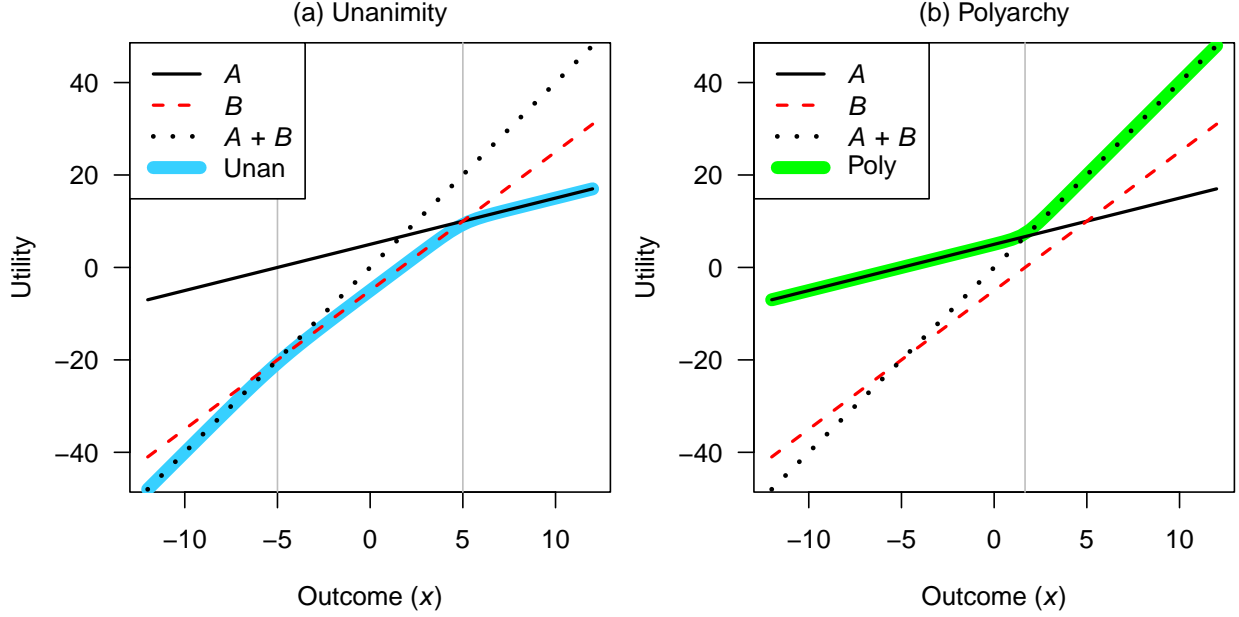


Figure 4: Utility functions for (a) unanimity and (b) polyarchy.

The utility function of unanimity we have uncovered is surprising for three reasons. First, it essentially corresponds to a simple mathematical function, the minimum of other utility functions. Second, the utility function of unanimity not only depends on the utility of  $A$  and the utility of  $B$ , but also on the utility of another individual that is not an actual member of the group, “individual  $A + B$ ,” whose utility function is the sum of the utility functions of  $A$  and  $B$  (the implication of this will be explored in the Results section). Finally, recall that marketing research had theorized that the utility function of a group was the average of the utility functions of its members. However, this clearly does not hold true for unanimity, and it will not be the case for polyarchy either, as we will observe next.

### 3.3 Polyarchy

We now repeat the same steps as above to derive the utility function of the polyarchy. The polyarchy approves a project when individual  $A$  OR  $B$  believes the project is a good one.

---

function and its approximation look very similar. For example, in Figure 4(a)  $u_{\text{Unan}}(4) = 6.87$  while  $\min(u_A(4), u_B(4), u_A(4) + u_B(4)) = u_B(4) = 7$ .

Hence, the screening function of the polyarchy is:<sup>4</sup>

$$\begin{aligned}
s_{\text{Poly}}(x) &= 1 - (1 - s_A(x))(1 - s_B(x)) \\
&= 1 - \left(1 - \frac{1}{1 + e^{-\alpha_A - \beta_A x}}\right) \left(1 - \frac{1}{1 + e^{-\alpha_B - \beta_B x}}\right) \\
&= 1 - \frac{1}{1 + e^{\alpha_A + \beta_A x}} \frac{1}{1 + e^{\alpha_B + \beta_B x}} \\
&= 1 - \frac{1}{1 + e^{\alpha_A + \beta_A x} + e^{\alpha_B + \beta_B x} + e^{\alpha_A + \beta_A x + \alpha_B + \beta_B x}}.
\end{aligned}$$

As before, we solve for the deterministic component of  $u_{\text{Poly}}(x)$  that produces this screening function:

$$\begin{aligned}
s(x) &= s_{\text{Poly}}(x) \\
\frac{1}{1 + e^{-u_{\text{Poly}}(x)}} &= 1 - \frac{1}{1 + e^{\alpha_A + \beta_A x} + e^{\alpha_B + \beta_B x} + e^{\alpha_A + \beta_A x + \alpha_B + \beta_B x}} \\
-1 + \frac{1}{1 + e^{-u_{\text{Poly}}(x)}} &= -\frac{1}{1 + e^{\alpha_A + \beta_A x} + e^{\alpha_B + \beta_B x} + e^{\alpha_A + \beta_A x + \alpha_B + \beta_B x}} \\
1 - \frac{1}{1 + e^{-u_{\text{Poly}}(x)}} &= \frac{1}{1 + e^{\alpha_A + \beta_A x} + e^{\alpha_B + \beta_B x} + e^{\alpha_A + \beta_A x + \alpha_B + \beta_B x}} \\
\frac{1}{1 + e^{u_{\text{Poly}}(x)}} &= \frac{1}{1 + e^{\alpha_A + \beta_A x} + e^{\alpha_B + \beta_B x} + e^{\alpha_A + \beta_A x + \alpha_B + \beta_B x}} \\
u_{\text{Poly}}(x) &= \log \left( e^{\alpha_A + \beta_A x} + e^{\alpha_B + \beta_B x} + e^{\alpha_A + \beta_A x + \alpha_B + \beta_B x} \right),
\end{aligned}$$

In other words, polyarchy does the opposite of unanimity to members' utilities. Panel (b) in Figure 4 illustrates this.

To summarize, Figure 2 retraces our approach and the derivations in this section: (a) starting from individual utility functions, (b) we compute the corresponding individual acceptance probabilities, (c) we then aggregate these probabilities, and (d) finally ask what would be the utility function that would produce these aggregate choice probabilities. The resulting utility function is the organizational utility function. The arrows connecting these four elements correspond to going back and forth between utility and probability space (arrows 1 and 3) and aggregating probabilities (arrow 2).

---

<sup>4</sup>The following derivations use the identity  $1 - 1/(1 + e^{-z}) = 1/(1 + e^z)$ .

## 4 Analyses

We now examine the behavior of the model under different scenarios. The goal is to develop deeper intuitions about how the behavior of unanimity and polyarchy depends on the types of individuals employed. We do so by developing four analyses: (i) comparing the risk-taking behavior of unanimity and polyarchy; (ii) increasing the number of organizational members; (iii) relaxing the “more is better” assumption to explore the effect of employing members with opposing interests (i.e., one utility function increases with  $x$  while the other decreases with  $x$ ); and (iv) examining how the results of classic strategic interaction models (Cournot competition and the principal–agent model) change when firms, instead of being risk-neutral, use a utility function corresponding to unanimity or polyarchy. In Appendix B, we illustrate the versatility of our approach by examining more complex individual utility functions.

### 4.1 Comparing the Risk-Taking Behavior of Unanimity and Polyarchy

So far we have shown that unanimity and polyarchy produce utility functions with very different shapes. In comparison with its members, unanimity amplifies negative utilities and dampens positive ones, while polyarchy does the opposite. These contrasting utility shapes have significant implications for risk assessment in project selection. To quantify these differences, we employ the standard concept of “certainty equivalent”—the guaranteed amount that would provide equal utility to a given risky proposition. This approach allows us to systematically compare how unanimity and polyarchy structures evaluate and respond to projects with uncertain outcomes.

We use the same individual utility functions and aggregation schemes (unanimity and polyarchy) as in Figure 4 to illustrate how aggregation can generate markedly different organizational risk preferences. Specifically, the individuals’ deterministic utility functions are  $u_A(x) = 5+x$  and  $u_B(x) = -5+3x$  (with the usual Logistic(0, 1) random-utility interpretation

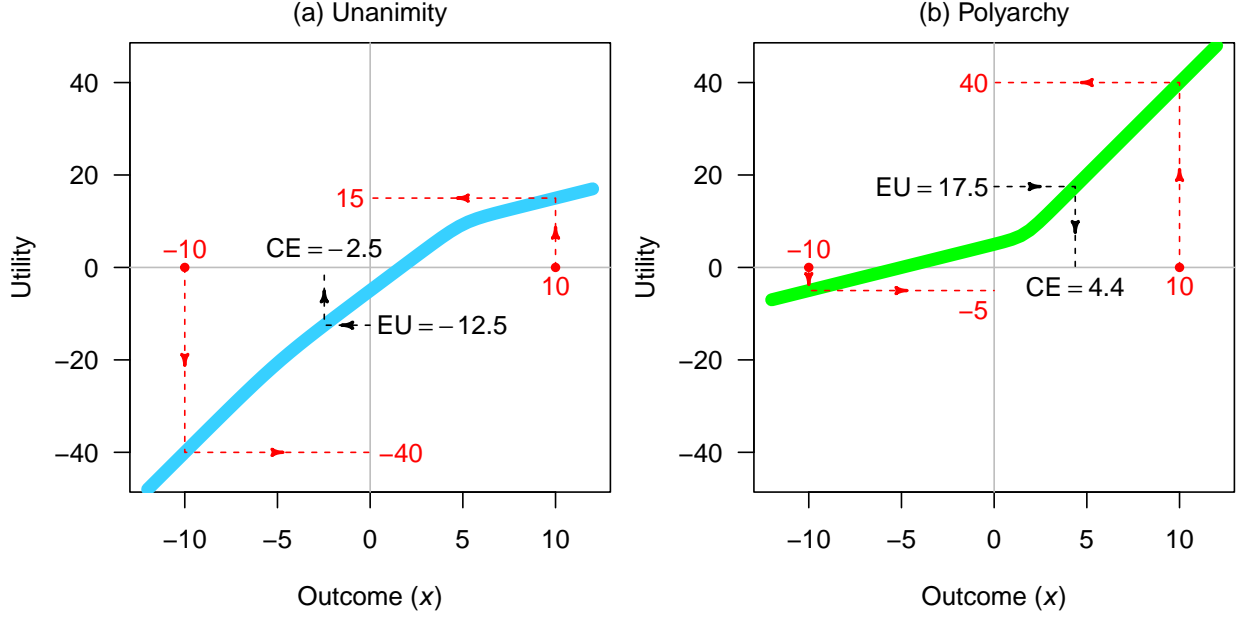


Figure 5: Certainty equivalent (CE) of a bet that pays either 10 or  $-10$  with probability 50% for unanimity and polyarchy.

in the background, as in Section 3). Figure 5 shows how unanimity and polyarchy evaluate a bet that pays either 10 or  $-10$  with equal probability. Under unanimity, the bet yields expected utility  $-12.5$  ( $= 0.5u_{\text{Unan}}(-10) + 0.5u_{\text{Unan}}(10) = 0.5 \times -40 + 0.5 \times 15$ ); under polyarchy, it yields expected utility  $17.5$  ( $= 0.5u_{\text{Poly}}(-10) + 0.5u_{\text{Poly}}(10) = 0.5 \times -5 + 0.5 \times 40$ ). Using certainty equivalents, unanimity treats the bet as equivalent to a sure loss of 2.5 ( $u_{\text{Unan}}(-2.5) = -12.5$ ), whereas polyarchy treats it as equivalent to a sure gain of 4.4 ( $u_{\text{Poly}}(4.4) = 17.5$ ). The implication is that polyarchy is far more likely to accept this bet than unanimity.<sup>5</sup>

A further illustration of the substantial difference in risk-aversion between unanimity and polyarchy is to think about what would be the minimum winning probability at which they would accept the bet (i.e.,  $p$  such that  $p \times u(10) + (1 - p) \times u(-10) = 0$ ). For the utility functions in Figure 5, unanimity would only accept the bet if the probability of winning was at least 73%, while polyarchy would just require an 11% probability of winning. In sum,

<sup>5</sup>We say “most likely” as these numerical examples did not incorporate the random utility component; however, the probability that Logistic(0, 1) noise would flip the sign of either expected utility is negligible in the example.

aggregation structures have a paramount effect on risk-taking behavior.<sup>6</sup>

The pronounced difference in risk aversion between unanimity and polyarchy, illustrated by their certainty equivalents for the same bet, highlights how aggregation structures shape organizational risk preferences derived from individual utilities. This analytical insight can inform our understanding of strategic decisions involving risk, such as corporate acquisitions. While studies often link acquisition choices to CEO attributes (Zhu and Chen 2015) or incentives (Wright et al. 2007), major acquisitions typically involve top management teams with heterogeneous risk preferences. Our framework suggests that aggregating these diverse preferences through structures like unanimity (more cautious) versus polyarchy (more risk-seeking) could help explain variations in acquisition aggressiveness, thus connecting individual managerial attributes to organizational outcomes (Finkelstein et al. 2009:115).

## 4.2 Effect of Increasing the Number of Members in the Organization

Using our method, one can derive the organizational utility functions given arbitrary individual utility functions and aggregation structures. Here we explore the case of unanimity and polyarchy as the number of members increases.<sup>7</sup>

Figure 6 illustrates the utility function of unanimity and polyarchy as the number of members ( $N$ ) increases. For simplicity, we assume that each member  $i$  has utility function  $U_i(x) = x + \varepsilon_i$ .

The main observation from this figure is that, as  $N$  increases, unanimity becomes more negative only for projects with negative  $x$ 's, while polyarchy becomes more positive only

---

<sup>6</sup>While scaling the utility functions matters, it does not alter the fundamental finding that unanimity is more risk-averse and polyarchy more risk-seeking. For instance, if the shape of the utility functions in Figure 5 remained the same but the  $y$ -axis range was compressed (e.g., from  $-4$  to  $+4$ ), polyarchy would still be more likely than unanimity to accept risky bets. Of course, if the range were minimal (e.g.,  $-0.04$  to  $+0.04$ ), noise would dominate, leading both structures to approve bets almost at random.

<sup>7</sup>To compute the formulas, one can redo the analyses in Section 3 using larger groups. But a simpler way of doing so is to rely on the fact that AND and OR are associative functions. For example, to derive the utility function of unanimity among three individuals ( $A$ ,  $B$ , and  $C$ ), one can compute the utility of  $A$  AND  $B$  and call that  $AB$ , and then compute the utility of  $AB$  AND  $C$ .

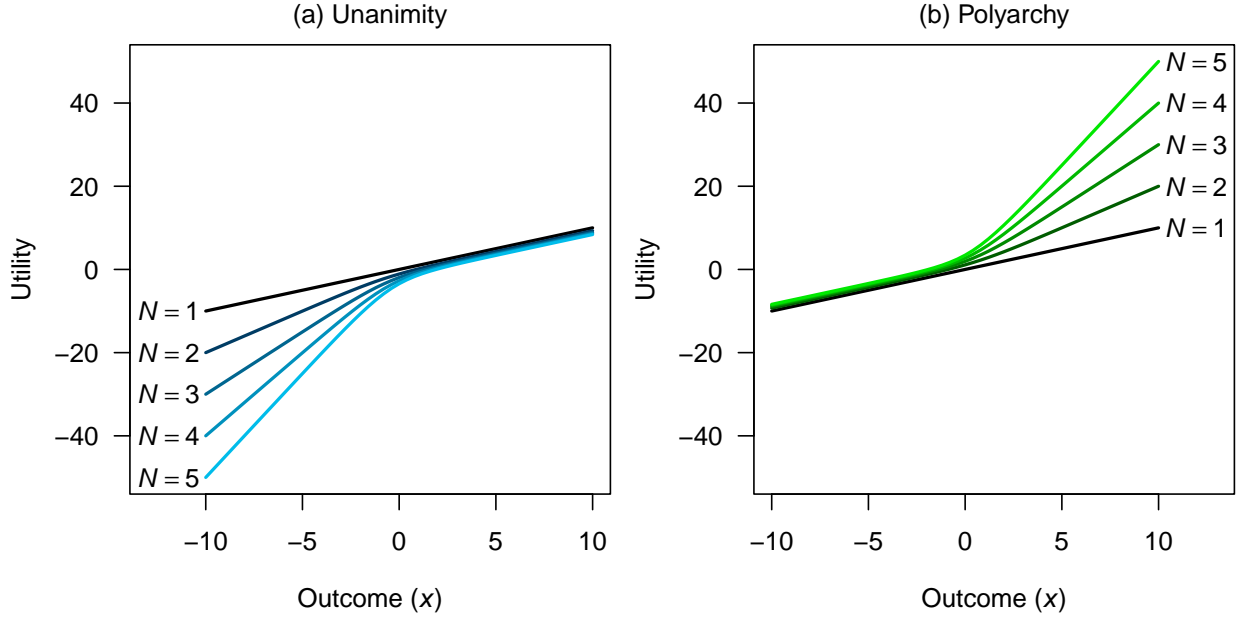


Figure 6: Effect of varying the number of individuals ( $N$ ) for unanimity and polyarchy. The figure assumes all individuals have utility function  $U_i(x) = x + \varepsilon_i$ .

for projects with positive  $x$ 's. This implies that unanimity becomes more risk-averse and polyarchy more risk-seeking as the number of members increases. To understand what drives these results, it is useful to walk through the case of unanimity among three members (i.e., the  $N = 3$  line in Figure 6(a)). The key is to realize that unanimity among three individuals is the same as unanimity between one individual and a group of two individuals who use unanimity (i.e., aggregating the  $N = 1$  and  $N = 2$  lines in Figure 6(a)). Doing so with the  $N = 1$  and  $N = 2$  lines produces the  $N = 3$  line. The same logic can be repeated for larger  $N$ 's as well as for polyarchy.

This analysis demonstrates how the derived organizational utility function, and consequently its risk preference, systematically shifts with team size ( $N$ ) and aggregation structure. Specifically, increasing  $N$  magnifies risk aversion under unanimity and risk-seeking under polyarchy. These predictable patterns offer clear avenues for empirically testing the link between team size, aggregation structure, and observed organizational risk-taking behavior.

### 4.3 Effect of Opposing Views within the Organization

So far we have studied individual utility functions that are increasing in  $x$ —although individuals may have different utility functions, they all agree that more of  $x$  is better. This is a reasonable assumption in many cases, such as when  $x$  represents a generally desirable attribute (e.g., profits, sales, or “likes”). Individuals may disagree on the utility of a given project and on whether a given project is worth accepting, but they agree that increasing  $x$  is a good thing.

However, there are cases where the “more is better” assumption is not appropriate. For example, if  $x$  represents a trait of a possible hire or where to locate. In cases like these, disagreements could run deep (e.g., members could disagree on whether to hire a younger or an older manager or on whether to locate close or far away from downtown). Here we analyze the effect of using unanimity and polyarchy with two individuals, one who likes and another who dislikes increasing  $x$ . Figure 7 illustrates such a case (the individuals’ utility functions used are  $U_A(x) = 1 + x + \varepsilon_A$  and  $U_B(x) = -1 - 0.5x + \varepsilon_B$ ). Again, we use simple linear utility functions for illustration; however, in this case, the slopes of the individuals are of a different sign to reflect opposite preferences.

Panel (a) in Figure 7 shows the utility of unanimity under these managers with opposing views. The organizational utility function in such a case follows an inverted-U shape. As before, the organizational utility function traces the minima of the underlying utility functions. In the case depicted in panel (a), this implies that on the negative side, the organizational utility function picks the upward-sloping utility function of  $A$  and, on the positive side, the downward-sloping utility function of  $B$ . Panel (b) illustrates the same individuals, now aggregated using polyarchy. In this case, the organizational utility function is now U-shaped.

Panels (c) and (d) in Figure 7 show the screening functions corresponding to the organizational utility functions from panels (a) and (b), respectively. An interesting observation here is that these screening functions look very different than the S-shaped screening functions

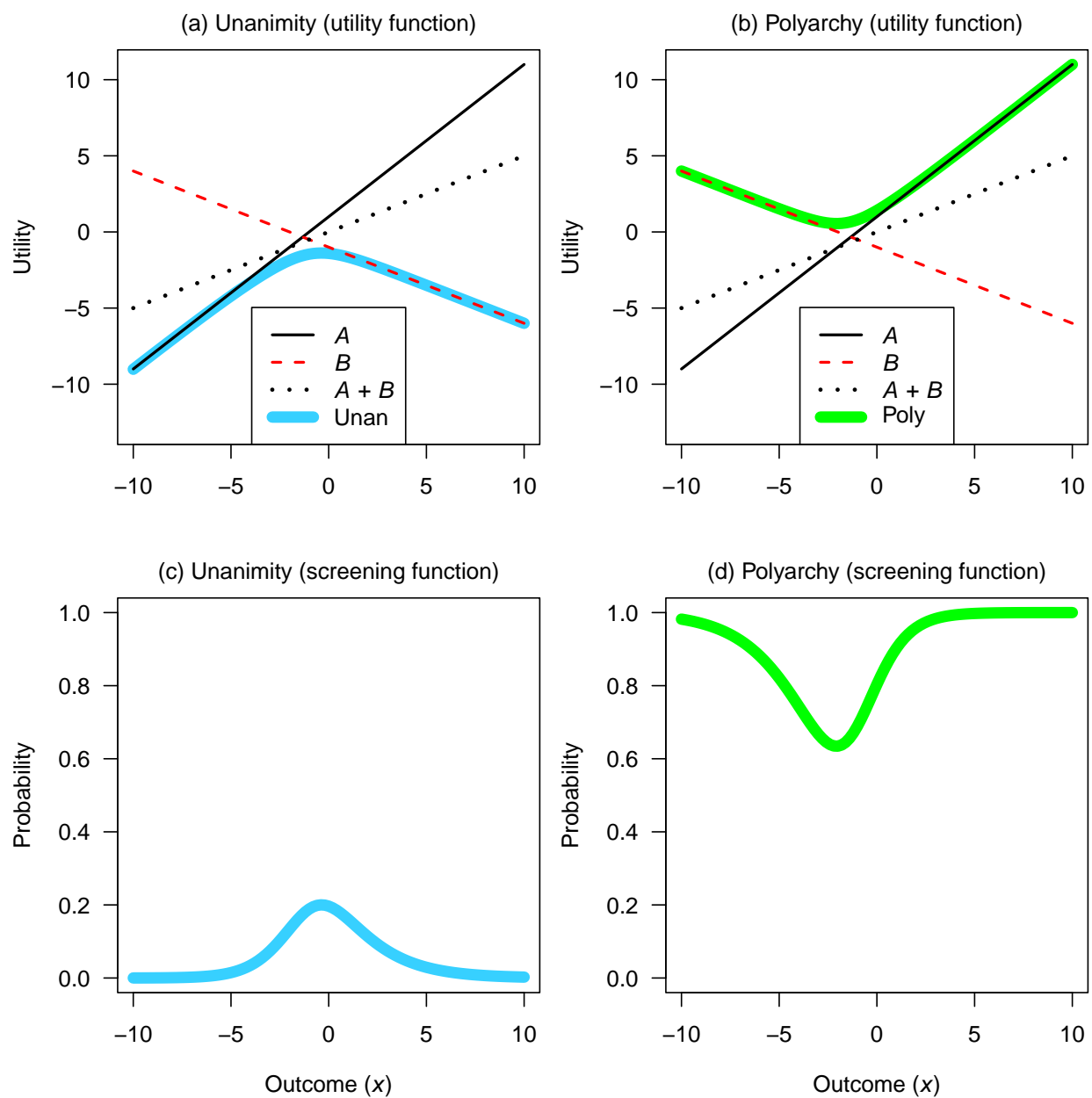


Figure 7: Utility and screening functions of unanimity and polyarchy when the individuals have opposing views. The thick lines in panels (a) and (b) show organizational utility functions and panels (c) and (d) show the respective organizational screening functions.

depicted earlier in this paper (Figures 1 and 3). One can understand this qualitative difference by following how the utility function affects the screening function. For an increasing utility function that goes from very negative to very positive utilities, its screening function will shift from mostly rejecting to mostly approving projects and, hence, will produce an S-shaped screening function like the ones in Figure 1. In contrast, if a utility function follows an inverted U-shape (as in panel (a) in Figure 7), its corresponding screening function will be most likely to accept projects at the apex of the inverted U-shape, which produces a bell-shaped screening function like the one shown in panel (c). The opposite happens for a polyarchy employing opposing individuals, which produces a U-shaped utility function and whose screening function achieves a minimum coinciding with the U-shape's nadir (i.e., panels (b) and (d) in Figure 7).

Panel (c) sheds light on situations characterized by deep disagreements. For instance, one could imagine this panel representing an academic department with two factions: a group that likes hiring theoreticians and a group that likes hiring empiricists (or in a corporate setting, this could represent a top management team divided between those prioritizing operations and those prioritizing marketing). The  $x$ -axis in this case represents the spectrum from pure theoreticians to pure empiricists. If hiring required unanimity, such a department would act as if its utility function followed an inverted U-shape. That is, only candidates who are neither too theoretical nor too empirical would be somewhat appealing to the department (as in panel (a)). Hence, this department would only accept candidates close to the middle of the spectrum. Moreover, because no candidate appeals across the board, the organizational utility is never very high and, consequently, this department may not hire very often (e.g., the screening function in panel (c) approves the most appealing candidate around 20% of the time). Of course, the particulars of conflict for different organizations and decisions will be different from those illustrated in Figure 7, but considering how an organization's utility function may look can help predict what the most likely organizational outcomes might be.

This formal approach to aggregation sheds light on how organizations manage the internal

conflict central to the Behavioral Theory of the Firm (Cyert and March 1963). Our model demonstrates *how* conflicting preferences among individuals or coalitions can be reconciled—or “quasi-resolved”—into organizational action based on the aggregation structure. As the preceding analysis of opposing views illustrates (Figure 7), unanimity forces compromise, favoring outcomes only where conflicting preferences find common ground (akin to a negotiated truce). Conversely, polyarchy allows the preferences of any single actor favoring action to drive the outcome, even amidst dissent. This provides a clear mechanism for understanding how seemingly coherent organizational goals can emerge despite, and indeed *as a result of*, aggregating conflicting internal viewpoints.

#### **4.4 Organizational Utility in Strategic Interactions: Applications to Collaboration and Competition**

We now illustrate the strategic implications of using these derived organizational utility functions by applying them to two classic interaction models, substituting the typical risk-neutral assumption with the distinct utility functions for unanimity and polyarchy developed earlier. First, we examine a Cournot model, analyzing how firms’ internal aggregation structures shape competitive output decisions. Second, we analyze a principal–agent model, representative of collaborative relationships like firm–supplier contracting, exploring how the principal’s aggregation structure influences optimal incentive design. These applications highlight how replacing the simplistic risk-neutral unitary actor assumption with behaviorally grounded organizational utility functions leads to richer predictions about strategic behavior in both competitive and collaborative settings (for complete technical derivations and parameter specifications, see Appendix C).

First, consider a Cournot duopoly where firms simultaneously choose production quantities under demand uncertainty. We compare all possible competitive scenarios among unanimity, polyarchy, and risk-neutral firms (labeled as  $U$ ,  $P$ , and  $N$  in Figure 8(a)). The game features

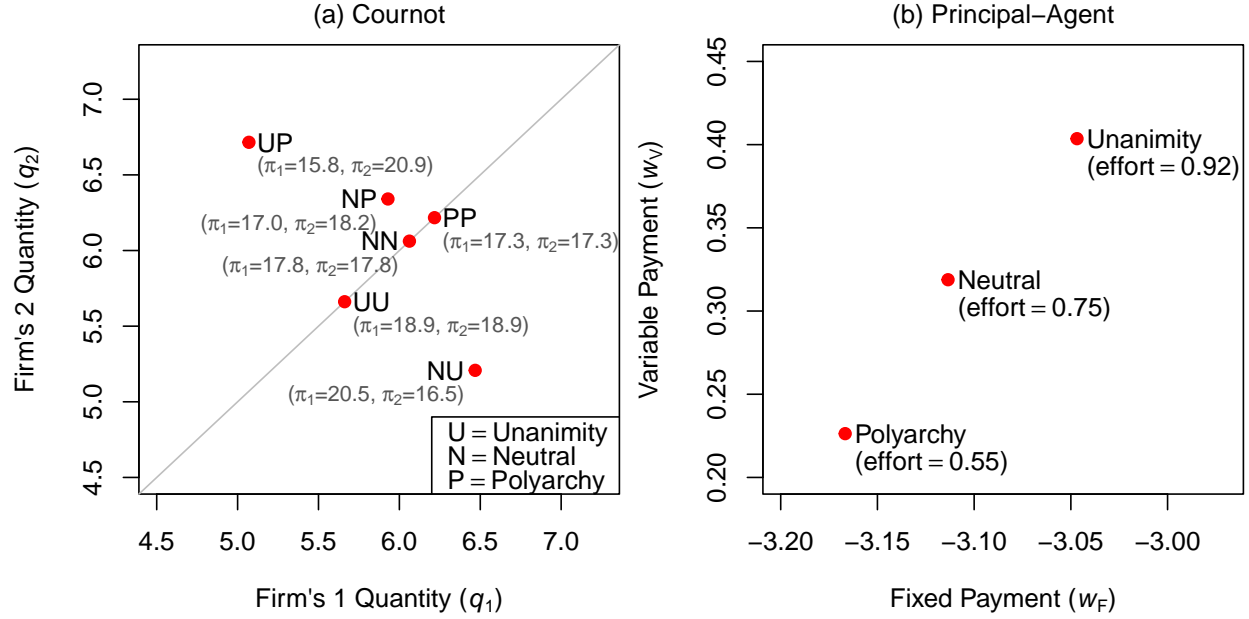


Figure 8: Equilibrium outcomes under different decision-making structures. (a) Production levels in Cournot competition between firms with different decision-making structures. (b) Principal-agent contracts and resulting effort levels under different principal decision-making structures.

a demand function with uncertainty in its intercept.<sup>8</sup> As Figure 8(a) shows, production is highest when two polyarchies compete (point  $PP$ ), because polyarchy's risk-seeking nature sees its utility increase the most with increased production. This implies that polyarchical firms in oligopolistic markets will tend to overproduce, whereas unanimity-based firms will tend to produce less. Interestingly, while firms would achieve higher profits by being more risk-averse and restricting production (point  $UU$ ), competitive pressures could drive them toward more risk-seeking structures, which increase output but reduce profitability.

Second, we examine a classic principal-agent relationship, modeling, for instance, a firm contracting with a supplier or employee. We assume a standard risk-averse agent<sup>9</sup> The principal sets a linear contract ( $wage = w_F + w_V \text{output}$ ). Figure 8(b) shows that a unanimity principal sets the strongest incentives (higher fixed and variable payments), thereby inducing maximum effort. This occurs because more risk-averse principals experience greater utility

<sup>8</sup>In this example,  $demand = intercept - 0.5(q_1 + q_2)$ , where  $intercept \sim \text{Normal}(10, 2^2)$ .

<sup>9</sup>We use a CARA utility function  $u_A = -\exp(-0.5 \cdot wage) - \frac{1}{2} \text{effort}^2$  and reservation utility of  $-5$ , and assume  $output \sim \text{Normal}(\text{effort}, 3^2)$ .

loss from low outputs and consequently set stronger incentives to prevent such outcomes. Less risk-averse principals do not prioritize such high effort as they are unconcerned with low output. Ultimately, risk-averse unanimities will tend to pay their agents (e.g., suppliers) higher rates to avoid the risk of low production. In contrast, risk-seeking polyarchies will tend to pay their agents less in order to optimize profitability while accepting the risk of reduced production.

These examples show how replacing the risk-neutral unitary actor assumption with realistic organizational utility functions offers different testable predictions about how individual preferences and aggregation structures translate into observable strategic behavior.

## 5 Discussion

Traditional strategy theories often depict firms as unitary actors, simplifying the complex interplay of individual preferences within an organization and assuming that the whole firm can be represented as having one utility function. The behavioral theory of the firm and the literature on microfoundations highlight that the unitary actor is an illusion: firms are composed of multiple actors with oftentimes conflicting goals, suggesting that the whole firm should not be understood as a unitary actor.

However, we show that if one knows the utility functions of the members of the organization and the aggregation structure they use, it is possible to derive an organizational utility function—one utility function that accurately represents the preferences of the organization. Thus, although the organization is not a unitary actor, if one uses the right organizational utility function, it is possible to analyze the organization *as if* it was a unitary actor. So, while the firm may be a motley crew behind the scenes, it can still put on a convincing one-person show for observers.

This approach bridges traditional strategy theories with a deeper understanding of organizational behavior. It acknowledges the complexity of internal dynamics while offering a

practical analytical framework. By recognizing both organizational intricacy and the potential for unified representation, we pave the way for more accurate and insightful strategy research. The following discussion explores the implications of our findings.

## 5.1 Contributions to the Strategy Literature

First, we show that organizational utility functions generally differ from the average of their members' utility functions, and the type of aggregation structure used greatly affects the shape of the resulting organizational utility function.<sup>10</sup> Specifically, unanimity traces the minima of underlying utility functions, while polyarchy traces their maxima. Interestingly, these underlying utility functions are not just the utility functions of the individual members of the organization but also include the utility functions of other, “synthetic” members that often amplify the characteristics of the actual members of the organization.

Second, our work provides a theoretical basis for revisiting unitary actor models to incorporate the role of individual heterogeneity and aggregation structures. We show that it is valid to conceptualize an organization in terms of a utility function as long as such a function is chosen carefully. This allows existing theories to be “retrofitted” by replacing their utility functions with ones that better reflect the characteristics of aggregation processes. For instance, search and industrial organization models commonly used in strategy assume risk-neutral firms (in terms of our model, such research assumes the organization's utility function is a straight line; Rothschild and Stiglitz 1976:631). However, our work shows that aggregation can exacerbate the characteristics of underlying utility functions; thus, most organizational utility functions, which arise from aggregation processes, may look quite different from those associated with unitary actor models so far. Retrofitting previous unitary actor theories offers a straightforward path for incorporating the role of individual heterogeneity and aggregation structures into these theories.

---

<sup>10</sup>Because the space of possible aggregation structures is vast (see, e.g., Csaszar 2013, Csaszar and Eggers 2013), it is likely that some aggregation structure could effectively average its members' utility functions—what we have pointed out is that this is not generally the case.

Third, we provide a way to address the call to better incorporate conflict into the behavioral theory of the firm (Gavetti et al. 2007). Our framework demonstrates how specific aggregation structures systematically combine potentially conflicting individual or coalition preferences (see, e.g., the predictions in the context of Figure 7). This process yields a distinct organizational utility function, offering a path to reconcile the apparent contradiction between viewing the organization as a coalition of diverse interests and ascribing it coherent goals (Cyert and March 1963:27). Moreover, by deriving this function analytically, our approach provides a tractable method for representing internal conflict’s impact on organizational preferences, tackling a long-standing challenge (March 1962:668–671).

Finally, our work can help the top management teams (TMT) literature establish a better link between the causes and effects it studies. Finkelstein et al. (2009:115) observation that “the gulf between executive characteristics and organizational outcomes is huge” is a call to better understand the chain of causality studied by this literature, which spans from individual characteristics (e.g., narcissism and functional background) to organizational outcomes (e.g., R&D investments and profitability). If such a chain of causality can be expressed in terms of individual utility functions, aggregation mechanisms, and organizational utility functions—and we think it can, as TMTs operate at the intersection of these elements—our work can help bridge the gulf at the core of the TMT literature. Hence, our work also suggests new, testable predictions for this literature regarding the effect of aggregation structures.

## **5.2 Limitations and Further Work**

As with all research, our work has limitations, which provide opportunities for future research. We identify four main areas for extension.

First, our analysis used stylized aggregation structures (unanimity, polyarchy) and simple individual utility functions. Future work could incorporate a wider array of aggregation mechanisms, such as majority voting, delegation, or weighted schemes representing power differentials (e.g., giving veto power or extra weight to certain individuals like a dictatorial

CEO; Csaszar 2013, Csaszar and Eggers 2013). Further extensions could model more complex individual utilities, perhaps incorporating multiple attributes, and explicitly distinguish between conflicts among individuals (Eisenhardt et al. 1997) and conflicts among goals (Gaba and Greve 2019, Obloj and Sengul 2020), potentially modeling the latter via different ways of aggregating multi-attribute utility functions.

Second, our model is static. Incorporating dynamics—how individual utilities or aggregation structures evolve through learning (Piezunka et al. 2022), strategic adaptation (Piezunka and Schilke 2023), or other decision processes—would provide richer insights into organizational change and adaptation. This could help illuminate phenomena like shifts in TMT decision patterns (Finkelstein et al. 2009) and incumbent responses to technological change (Eggers and Park 2018); for example, how the interplay between management team composition, individual utilities, and the aggregation mechanism shapes the firm’s overall risk preference, influencing decisions about whether to invest in or actively capture value from emerging technologies (Christensen and Bower 1996).

Third, regarding empirical work, future research could test our model in the lab by eliciting the utility functions of group members and checking whether the utility function of the group matches our predictions. Apart from validating our work, this research would shed light on the effect size of aggregation structure vis-à-vis other individual-, firm-, and environmental-level contingencies.

Finally, our framework naturally extends to hybrid decision-making bodies where humans and AI agents jointly evaluate opportunities. By treating an AI system as another organizational “member” with its own utility function and error profile, researchers can analyze human–AI collaboration using the same primitives as human-only teams. Like human members, an AI’s utility function is defined by its training and constraints, meaning it may be misaligned with organizational goals or opaque to designers. Consequently, firms must validate what the AI is actually optimizing behaviorally rather than assuming it is a neutral oracle. The key design implication is that firms must jointly choose the AI’s utility function

and its structural role—such as a gatekeeper (unanimity) or a spotter (polyarchy)—as these choices fundamentally reshape the organization’s risk profile. This perspective opens a new research agenda focused on characterizing AI utility functions and designing aggregation structures that optimally aggregate human and machine recommendations.

These extensions would improve our understanding of how micro-level preferences and power structures translate into macro-level organizational behavior and strategic decision-making.

### **5.3 Conclusion**

Our paper provides insight into how heterogeneous individual utility functions are aggregated by different decision-making structures. In doing so, we demonstrate that organizational utility functions are a meaningful construct and that they are not just an average of individuals’ utility functions. Rather, part of an organization’s utility function is directly inherited from its members’ utility functions, but part of it also comes from adding up individuals’ utility functions. How these parts are combined critically depends on the aggregation structure used by the organization. The picture that emerges is that organizations can exaggerate the behavioral characteristics of their members. It is the role of the organization designer to harness these exaggerated traits in ways that benefit the organization. Ultimately, organizational utility functions can help provide keener insights into how organizations make the decisions they do. Overall, our goal has been to logically connect key micro and macro aspects of organizations and, thus, to help provide a sound foundation to understand the behavior of organizations based on the attributes of their underlying individuals.

## References

- Adamowicz, W., M. Hanemann, J. Swait, R. Johnson, D. Layton, M. Regenwetter, T. Reimer, R. Sorkin. 2005. Decision strategy and structure in households: A ‘groups’ perspective. *Marketing Letters* **16**(3–4) 387–399.
- Aime, F., S. Johnson, J. W. Ridge, A. D. Hill. 2009. The routine may be stable but the advantage is not: Competitive implications of key employee mobility. *Strategic Management Journal* **31**(1) 75–87.
- Arora, N., G. M. Allenby. 1999. Measuring the influence of individual preference structures in group decision making. *Journal of Marketing Research* **36**(4) 476.
- Arrow, K. J. 1951. *Social Choice and Individual Values*. Yale University Press, New Haven, CT.
- Barney, J., T. Felin. 2013. What are microfoundations? *Academy of Management Perspectives* **27**(2) 138–155.
- Bingham, C. B., T. Howell, T. E. Ott. 2019. Capability creation: Heuristics as microfoundations. *Strategic Entrepreneurship Journal* **13**(2) 121–153.
- Binmore, K. 2009. Interpersonal comparison of utility. D. Ross, H. Kincaid, eds., *The Oxford Handbook of Philosophy of Economics*, chap. 20. Oxford University Press, New York, 540–559.
- Black, D. 1958. *The Theory of Committees and Elections*. Cambridge University Press, Cambridge, UK.
- Boyd, S. P., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, Cambridge, UK.
- Christensen, C. M., J. L. Bower. 1996. Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal* **17**(3) 197–218.
- Christensen, M., T. Knudsen. 2010. Design of decision making-organizations. *Management Science* **56**(1) 71–89.
- Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* **5**(4) 559–583.
- Coase, R. H. 1937. The nature of the firm. *Economica* **4**(16) 386–405.
- Coase, R. H. 1988. The nature of the firm: Origin. *Journal of Law, Economics, & Organization* **4**(1) 3–17.
- Condorcet, J. A. N. 1785/1994. *Foundations of Social Choice and Political Theory*. E. Elgar, Brookfield, VT. Translated and edited by I. McLean and F. Hewitt.
- Cournot, A. A. 1838/1971. *Researches into the Mathematical Principles of the Theory of Wealth*. A. M. Kelley, New York.
- Cowen, A. P., F. Rink, I. R. P. Cuypers, D. A. Grégoire, I. Weller. 2022. Applying Coleman’s boat in management research: Opportunities and challenges in bridging macro and micro theory. *Academy of Management Journal* **65**(1) 1–10.
- Csaszar, F. A. 2012. Organizational structure as a determinant of performance: Evidence from mutual funds. *Strategic Management Journal* **33**(6) 611–632.
- Csaszar, F. A. 2013. An efficient frontier in organization design: Organizational structure as a determinant of exploration and exploitation. *Organization Science* **24**(4) 1083–1101.
- Csaszar, F. A., J. P. Eggers. 2013. Organizational decision making: An information aggregation view. *Management Science* **59**(10) 2257–2277.

- Cyert, R. M., J. G. March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall, Englewood Cliffs, NJ.
- Davis, J. H. 1973. Group decision and social interaction: Theory of social decision schemes. *Psychological Review* **80**(2) 97–125.
- Eggers, J. P., K. F. Park. 2018. Incumbent adaptation to technological change: The past, present, and future of research on heterogeneous incumbent response. *Academy of Management Annals* **12**(1) 357–389.
- Eisenhardt, K. M., J. L. Kahwajy, L. J. Bourgeois. 1997. Conflict and strategic choice: How top management teams disagree. *California Management Review* **39**(2) 42–62.
- Englmaier, F., N. J. Foss, T. Knudsen, T. Kretschmer. 2019. Organization design and firm heterogeneity: Towards an integrated research agenda for strategy. *Advances in Strategic Management* **40** 229–252.
- Fechner, G. T. 1860. *Elemente der Psychophysik*. Breitkopf und Härtel, Leipzig.
- Felin, T., N. J. Foss, R. E. Ployhart. 2015. The microfoundations movement in strategy and organization theory. *Academy of Management Annals* **9**(1) 575–632.
- Finkelstein, S., D. C. Hambrick, A. A. Cannella. 2009. *Strategic Leadership: Theory and Research on Executives, Top Management Teams, and Boards*. Oxford University Press, New York.
- Friedkin, N. E., E. C. Johnsen. 2011. *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge University Press, New York.
- Gaba, V., H. R. Greve. 2019. Safe or profitable? The pursuit of conflicting goals. *Organization Science* **30**(4) 647–667.
- Gavetti, G., D. A. Levinthal, W. Ocasio. 2007. Neo-Carnegie: The Carnegie School's past, present, and reconstructing for the future. *Organization Science* **18**(3) 523–536.
- Grossman, S. J., O. D. Hart. 1986. The costs and benefits of ownership: A theory of vertical and lateral integration. *The Journal of Political Economy* **94**(4) 691–719.
- Haleblian, J., C. E. Devers, G. McNamara, M. A. Carpenter, R. B. Davison. 2009. Taking stock of what we know about mergers and acquisitions: A review and research agenda. *Journal of Management* **35**(3) 469–502.
- Hambrick, D. C., P. A. Mason. 1984. Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review* **9**(2) 193–206.
- Hammond, P. J. 1991. Interpersonal comparisons of utility: Why and how they are and should be made. J. Elster, J. E. Roemer, eds., *Interpersonal Comparisons of Well-Being*, chap. 7. Cambridge University Press, Cambridge, UK, 200–225.
- Hart, O. 1995. Corporate governance: Some theory and implications. *Economic Journal* **105**(430) 678–689.
- Hastie, R., T. Kameda. 2005. The robust beauty of majority rules in group decisions. *Psychological Review* **112**(2) 494–508.
- Helfat, C. E., M. A. Peteraf. 2015. Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal* **36**(6) 831–850.
- Hess, A. M., F. T. Rothaermel. 2011. When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry. *Strategic Management Journal* **32**(8) 895–909.
- Hess, S., A. J. Daly. 2014. *Handbook of Choice Modelling*. Edward Elgar, Cheltenham, UK.

- Hogarth, R. M. 1978. A note on aggregating opinions. *Organizational Behavior and Human Performance* **21**(1) 40–46.
- Holmstrom, B. 1979. Moral hazard and observability. *Bell Journal of Economics* **10** 74–91.
- Jensen, M. C., W. H. Meckling. 1976. Theory of firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* **3**(4) 305–360.
- Joseph, J., V. Gaba. 2020. Organizational structure, information processing, and decision-making: A retrospective and road map for research. *Academy of Management Annals* **14**(1) 267–302.
- Kahneman, D., O. Sibony, C. R. Sunstein. 2021. *Noise: A Flaw in Human Judgment*. Little, Brown Spark, New York.
- Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263–291.
- Knudsen, T., D. A. Levinthal. 2007. Two faces of search: Alternative generation and alternative evaluation. *Organization Science* **18**(1) 39–54.
- Leiblein, M. J., J. J. Reuer, T. R. Zenger. 2018. What makes a decision strategic? *Strategy Science* **3**(4) 558–573.
- Levinthal, D. A. 1997. Adaptation on rugged landscapes. *Management Science* **43**(7) 934–950.
- Lim, E. N. K., B. T. McCann. 2014. Performance feedback and firm risk taking: The moderating effects of CEO and outside director stock options. *Organization Science* **25**(1) 262–282.
- Loon, M., L. Otaye-Ebede, J. Stewart. 2020. Thriving in the new normal: The HR microfoundations of capabilities for business model innovation—an integrated literature review. *Journal of Management Studies* **57**(3) 698–726.
- Luce, R. D. 1959. *Individual Choice Behavior*. John Wiley and Sons, New York.
- Lugosch, L. 2020. A contemplation of logsumexp. Available at: <https://lorenlugosch.github.io/posts/2020/06/logsumexp/> (accessed March 28, 2025).
- Manski, C. F. 1977. The structure of random utility models. *Theory and Decision* **8**(3) 229–254.
- March, J. G. 1962. The business firm as a political coalition. *The Journal of Politics* **24**(4) 662–678.
- Marchand, A. 2014. Joint consumption challenges in groups. *Journal of Consumer Marketing* **31**(6/7) 483–493.
- Marschak, J., R. Radner. 1972. *Economic Theory of Teams*. Yale University Press, New Haven, CT.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed., *Frontiers in Econometrics*, chap. 4. Academic Press, New York, 105–142.
- Menasco, M. B., D. J. Curry. 1989. Utility and choice: An empirical study of wife/husband decision making. *Journal of Consumer Research* **16**(1) 87–97.
- Meyer-Doyle, P., S. Lee, C. E. Helfat. 2019. Disentangling the microfoundations of acquisition behavior and performance. *Strategic Management Journal* **40**(11) 1733–1756.
- Mithani, M. A., J. P. O’Brien. 2020. So what exactly is a ‘coalition’ within an organization? A review and organizing framework. *Journal of Management* **47**(1) 171–206.
- Nag, R., D. C. Hambrick, M. J. Chen. 2007. What is strategic management, really? Inductive derivation of a consensus definition of the field. *Strategic Management Journal* **28**(9) 935–955.
- Obloj, T., M. Sengul. 2020. What do multiple objectives really mean for performance? Empirical evidence from the French manufacturing sector. *Strategic Management Journal* **41**(13) 2518–2547.
- Piezunka, H., V. A. Aggarwal, H. E. Posen. 2022. The aggregation–learning trade-off. *Organization Science* **33**(3) 1094–1115.

- Piezunka, H., O. Schilke. 2023. The dual function of organizational structure: Aggregating and shaping individuals' votes. *Organization Science* **34**(5) 1914–1937.
- Porter, M. E. 1985. *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press, New York.
- Rothschild, M., J. Stiglitz. 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics* **90**(4) 629–649.
- Sah, R. K., J. E. Stiglitz. 1986. The architecture of economic systems: Hierarchies and polyarchies. *American Economic Review* **76**(4) 716–727.
- Sah, R. K., J. E. Stiglitz. 1988. Committees, hierarchies and polyarchies. *Economic Journal* **98**(391) 451–470.
- Soll, J. B., R. P. Larrick. 2009. Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **35**(3) 780–805.
- Thurstone, L. L. 1927. A law of comparative judgment. *Psychological Review* **34**(4) 273–286.
- Tirole, J. 1988. *The Theory of Industrial Organization*. MIT Press, Cambridge, MA.
- Tuschke, A., W. M. G. Sanders, E. Hernandez. 2013. Whose experience matters in the boardroom? The effects of experiential and vicarious learning on emerging market entry. *Strategic Management Journal* **35**(3) 398–418.
- Williamson, O. E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York.
- Wright, P., M. Kroll, J. A. Krug, M. Pettus. 2007. Influences of top management team incentives on firm risk taking. *Strategic Management Journal* **28**(1) 81–89.
- Zhang, C. M., H. R. Greve. 2019. Dominant coalitions directing acquisitions: Different decision makers, different decisions. *Academy of Management Journal* **62**(1) 44–65.
- Zhu, D. H., G. Chen. 2015. Narcissism, director selection, and risk-taking spending: Narcissism, director selection, and risk-taking spending. *Strategic Management Journal* **36**(13) 2075–2098.

# Appendix A: The LogSumExp Approximation

This appendix provides a detailed explanation of the LogSumExp approximation, which offers a convenient intuition for interpreting the organizational utility functions derived in the main text.

## A.1 The LogSumExp Function and Its Bounds

The LogSumExp function is defined as  $\text{LSE}(x_1, \dots, x_n) = \log(e^{x_1} + e^{x_2} + \dots + e^{x_n})$ . This function appears frequently in machine learning, convex optimization, and statistical mechanics (see, e.g., Boyd and Vandenberghe 2004:72, Lugosch 2020). A key property of this function is that it closely approximates the maximum function, with the following tight bounds:

$$\max(x_1, x_2, \dots, x_n) \leq \log(e^{x_1} + e^{x_2} + \dots + e^{x_n}) \leq \max(x_1, x_2, \dots, x_n) + \log n$$

### A.1.1 Proof

Let  $M = \max(x_1, x_2, \dots, x_n)$ .

**Lower Bound.** Since  $M$  is the maximum,  $e^M$  is the largest term in the sum  $\sum_{i=1}^n e^{x_i}$ . Thus,  $\sum_{i=1}^n e^{x_i} \geq e^M$ . Taking the logarithm (a monotonically increasing function) preserves the inequality:

$$\log\left(\sum_{i=1}^n e^{x_i}\right) \geq \log(e^M) = M.$$

**Upper Bound.** For all  $i$ , we have  $x_i \leq M$ , so  $e^{x_i} \leq e^M$ . Summing over all  $i$ :

$$\sum_{i=1}^n e^{x_i} \leq \sum_{i=1}^n e^M = ne^M.$$

Taking the logarithm:

$$\log\left(\sum_{i=1}^n e^{x_i}\right) \leq \log(ne^M) = \log n + M.$$

**Conclusion.** Combining both bounds:

$$M \leq \log \left( \sum_{i=1}^n e^{x_i} \right) \leq M + \log n.$$

This shows that the LogSumExp function is tightly approximated by  $\max(x_1, x_2, \dots, x_n)$ . The approximation error is bounded above by  $\log n$ , which equals approximately 1.1 for  $n = 3$  (the number of terms in our two-member organizational utility functions). When the differences between the  $x_i$  values are large, the approximation becomes even tighter, as the sum becomes dominated by the largest exponential term.

## A.2 Application to Organizational Utility Functions

### A.2.1 Unanimity

The utility function for unanimity derived in the main text is:

$$u_{\text{Unan}}(x) = -\log(e^{-\alpha_A - \beta_A x} + e^{-\alpha_B - \beta_B x} + e^{-\alpha_A - \alpha_B - \beta_A x - \beta_B x}).$$

This can be rewritten as  $u_{\text{Unan}}(x) = -\text{LSE}(-\alpha_A - \beta_A x, -\alpha_B - \beta_B x, -\alpha_A - \alpha_B - \beta_A x - \beta_B x)$ .

Applying the LogSumExp approximation:

$$\begin{aligned} u_{\text{Unan}}(x) &\approx -\max(-\alpha_A - \beta_A x, -\alpha_B - \beta_B x, -\alpha_A - \alpha_B - \beta_A x - \beta_B x) \\ &= \min(\alpha_A + \beta_A x, \alpha_B + \beta_B x, \alpha_A + \alpha_B + \beta_A x + \beta_B x) \\ &= \min(u_A(x), u_B(x), u_A(x) + u_B(x)). \end{aligned}$$

Thus, the utility function of unanimity traces the pointwise minimum of the individual utility functions and their sum.

### A.2.2 Polyarchy

The utility function for polyarchy derived in the main text is:

$$u_{\text{Poly}}(x) = \log(e^{\alpha_A + \beta_A x} + e^{\alpha_B + \beta_B x} + e^{\alpha_A + \alpha_B + \beta_A x + \beta_B x}).$$

This is directly in LogSumExp form:  $u_{\text{Poly}}(x) = \text{LSE}(\alpha_A + \beta_A x, \alpha_B + \beta_B x, \alpha_A + \alpha_B + \beta_A x + \beta_B x)$ .

Applying the approximation:

$$\begin{aligned} u_{\text{Poly}}(x) &\approx \max(\alpha_A + \beta_A x, \alpha_B + \beta_B x, \alpha_A + \alpha_B + \beta_A x + \beta_B x) \\ &= \max(u_A(x), u_B(x), u_A(x) + u_B(x)). \end{aligned}$$

Thus, the utility function of polyarchy traces the pointwise maximum of the individual utility functions and their sum.

## Appendix B: More Complex Utility Functions

For the sake of clarity, so far we have used linear individual utility functions that depend on just one parameter ( $x$ ). But neither of these assumptions is required for our method to work: the steps described in Figure 2 can be used to aggregate arbitrarily complex utility functions. Below we illustrate this point by aggregating non-linear and multi-parameter utility functions. As in the main text, utilities are interpreted in random-utility form (i.e.,  $U(\cdot) = u(\cdot) + \varepsilon$  with  $\varepsilon \sim \text{Logistic}(0, 1)$ ); the figures below plot the deterministic components  $u(\cdot)$ .

Figure B.1 illustrates how unanimity and polyarchy aggregate two risk-averse individuals. The individuals in this example have exponential utility functions  $u_A(x) = 10(1 - e^{-x/10})$  and  $u_B(x) = 10(1 - e^{-x/5})$  and hence, exhibit constant absolute risk aversion. Aggregation operates the same as before: unanimity and polyarchy trace the pointwise minima and maxima of the utility functions of  $A$ ,  $B$ , and  $A + B$ . As before also, unanimity exacerbates risk-aversion and polyarchy, risk-seeking.

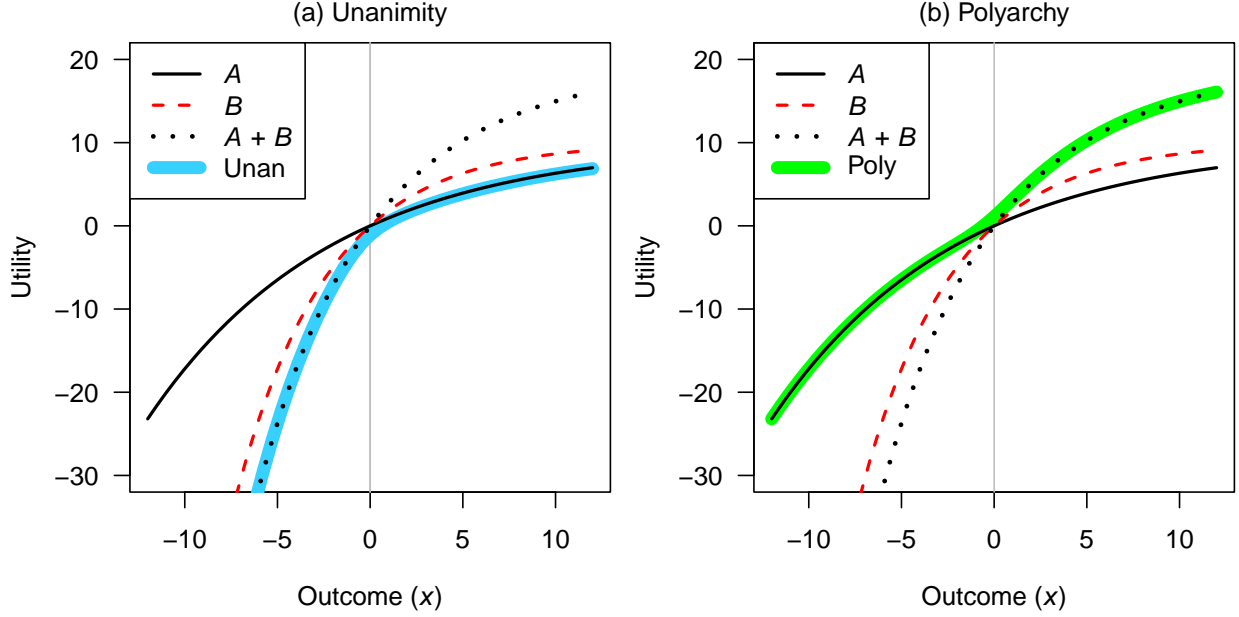


Figure B.1: Aggregating risk-averse individuals.

Figure B.2 illustrates the aggregation of utility functions that depend on two arguments ( $x_1$  and  $x_2$ ). For instance, these two variables could correspond to two characteristics of a project under consideration (e.g., the range and power of a car). In the example, the two individuals differ in how much they value each characteristic. Panel (a) plots the utility functions of both individuals ( $u_A(x_1, x_2) = x_1 + x_2$  and  $u_B(x_1, x_2) = 2x_1 + 3x_2$ ). Panel (b) shows the corresponding utility function of unanimity and polyarchy. Similarly to the previous results, the utility for polyarchy is always above that for unanimity. Also, the organizational utility rises sharply for positive values of  $x_1$  and  $x_2$  for polyarchy and drops steeply for negative values of  $x_1$  and  $x_2$  for unanimity.

## Appendix C: Details of Strategic Interaction Analyses

This appendix provides the technical details, functional forms, and parameter values required to reproduce the analyses presented in Section 4.4. Throughout this appendix, we model organizations composed of two members,  $A$  and  $B$ , with the same random-utility specifications used in Figure 4. Specifically,  $U_A(x) = 5 + x + \varepsilon_A$  and  $U_B(x) = -5 + 3x + \varepsilon_B$ , where the

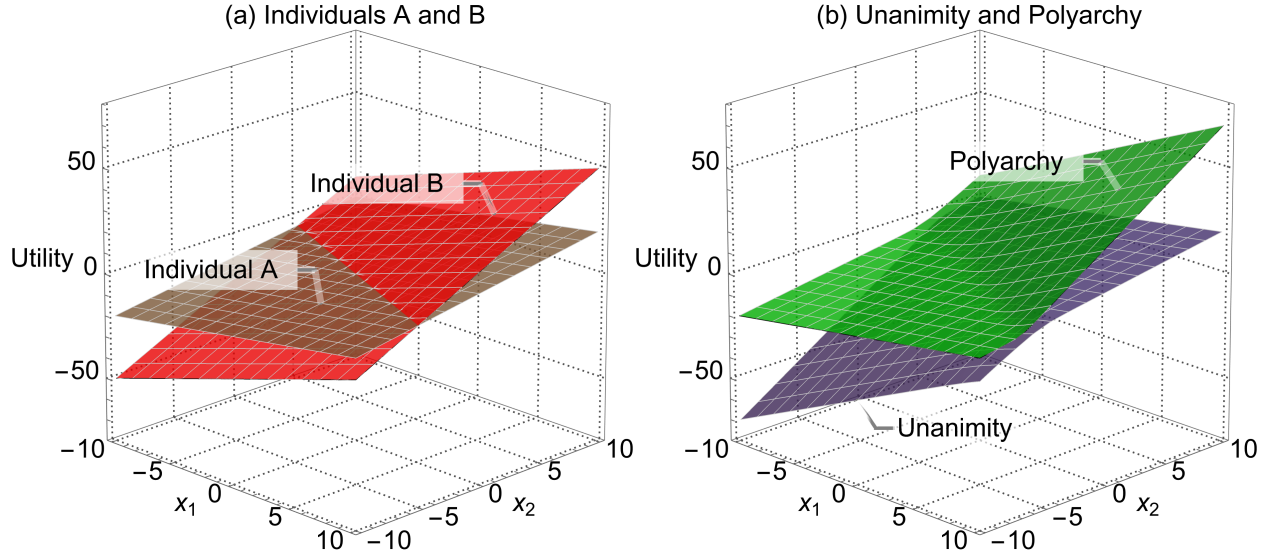


Figure B.2: Aggregating utility functions that depend on two arguments ( $x_1$  and  $x_2$ ).

error terms  $\varepsilon_A$  and  $\varepsilon_B$  are independent draws from a Logistic(0, 1) distribution.

## C.1 Cournot Competition Model

We model a duopoly where two firms ( $i$  and  $j$ ) simultaneously choose production quantities ( $q_i$  and  $q_j$ ) to maximize their expected utility.

*Market Structure.* We assume a standard linear inverse demand function with uncertainty in the intercept, where market price  $P$  depends on total quantity  $Q = q_i + q_j$ :

$$P(Q) = \max(a - b(q_i + q_j), 0)$$

where  $a$  is a random variable drawn from a Normal distribution  $a \sim \text{Normal}(\bar{a}, \sigma_a^2)$ . The marginal cost of production,  $c$ , is constant and identical for both firms. The profit for firm  $i$  is given by  $\pi_i = (P(Q) - c)q_i$ .

*Firm Preferences.* Firms evaluate profits according to their organizational utility function,  $u(\pi)$ . We compare three types of firms: risk-neutral (unitary actor), unanimity, and polyarchy. The risk-neutral firm employs a linear utility function  $u(\pi) = \pi$ . The unanimity and polyarchy firms aggregate the preferences of two members ( $A$  and  $B$ ) using the functional forms derived

in Section 3.

*Optimization and Equilibrium.* Each firm  $i$  chooses  $q_i$  to maximize its expected utility given the quantity produced by the competitor  $q_j$ . We solve for the Nash Equilibrium numerically using an iterative best-response algorithm where firms update their quantities sequentially until convergence (defined as a change in quantity of less than  $10^{-7}$  between iterations). Expected utility integrals are computed numerically over the range  $[\bar{a} - 10\sigma_a, \bar{a} + 10\sigma_a]$ .

*Parameters.* Table C.1 lists the market parameter values used to generate Figure 8(a).

Parameter	Symbol	Value
Mean demand intercept	$\bar{a}$	10
SD of demand intercept	$\sigma_a$	2
Demand slope	$b$	0.5
Marginal cost	$c$	1

Table C.1: Parameters for Cournot competition analysis.

## C.2 Principal–Agent Collaboration Model

We model a collaboration where a principal contracts a risk-averse Agent to perform a task.

*Production and Contract.* The output  $R$  is determined by the Agent’s effort  $e$  and a random shock, such that  $R \sim \text{Normal}(e, \sigma^2)$ . The principal offers a linear contract consisting of a fixed wage  $w_F$  and a variable component  $w_V$  based on output,  $w = w_F + w_V R$ .

*Agent Preferences.* The Agent is risk-averse with Constant Absolute Risk Aversion (CARA) utility and a convex cost of effort. The Agent’s utility function is:

$$U_{\text{Agent}}(w, e) = -e^{-\gamma w} - \frac{1}{2}e^2$$

where  $\gamma$  is the coefficient of absolute risk aversion.

*Principal Preferences.* The principal evaluates their net income  $(R - w)$  using one of the three organizational utility functions defined previously ( $u_{\text{Neutral}}$ ,  $u_{\text{Unan}}$ , or  $u_{\text{Poly}}$ ).

*Optimization Problem.* The principal chooses the contract terms  $(w_F, w_V)$  and the induced effort level  $e$  to maximize their expected utility, subject to the Agent’s Incentive Compatibility

(IC) and Participation Constraints (PC):

$$\begin{aligned}
 & \max_{w_F, w_V, e} \mathbb{E}_R[u_{\text{Principal}}(R - w)] \\
 & \text{s.t. } e = \arg \max_{e'} \mathbb{E}_R[U_{\text{Agent}}(w, e')] \quad (\text{IC}) \\
 & \mathbb{E}_R[U_{\text{Agent}}(w, e)] \geq \underline{U} \quad (\text{PC})
 \end{aligned}$$

The IC constraint is implemented via the Agent’s first-order condition with respect to effort. The problem is solved using numerical optimization (Simulated Annealing method).

*Parameters.* Table C.2 lists the task environment parameter values used to generate Figure 8(b).

Parameter	Symbol	Value
Output noise (SD)	$\sigma$	3
Agent risk aversion	$\gamma$	0.5
Agent reservation utility	$\underline{U}$	-5

Table C.2: Parameters for principal–agent analysis.