Voting and Experimentation

Bruno Strulovici*

December 29, 2007

Abstract

This paper analyzes a dynamic voting model where individual preferences evolve through experimentation. Individual votes reflect not only current preferences but also the anticipated effect of elected alternatives on future preferences and votes. The analysis is conducted in a two-arm bandit model, with a safe (status quo) alternative and a risky alternative whose payoff distribution, or "type", varies across individuals and may be learned through experimentation. Under any voting rule, society experiments less than any individual would if he could dictate future decisions, due to a *control-loss effect*. Depending on the nature of uncertainty, majority-based experimentation also has a systematic bias compared to the utilitarian policy. For *large* groups with independently distributed types, this control-loss effect annihilates the value of experimentation, prompting individuals to vote myopically. However, even with independently distributed types, a positive news shock for anyone raises everyone's value function and incentive to experiment. Efficiency increases with *ex ante* preference correlation. The paper also discusses the effect on experimentation of the ability to commit and of asymmetric information.

Keywords: Experimentation, Dynamic Social Choice, Reforms, Conservatism.

*Nuffield College and Economics Department, Oxford University. I am very grateful to Meg Meyer, Andrea Patacconi, and Kevin Roberts for insightful conversations, as well as to Daron Acemoglu, Patrick Bolton, Alessandra Casella, Robert Evans, Emeric Henry, Godfrey Keller, Jeremy Large, Rocco Macchiavello, Paul Milgrom, Roger Myerson, Volker Nocke, Andrea Prat, John Quah, Marzena Rostek, Gilles Serra and other seminar participants at Oxford University, Gerzensee's European Summer Symposium in Economic Theory, and Cambridge University for numerous comments. I am also particularly grateful to Paul Milgrom for his general insights and encouragements about research, which had a significant influence on this project.

1 Introduction

The positive social choice literature usually assumes that individuals perfectly know their own preferences. In reality, however, preferences may evolve through learning. For example, a reform may have uncertain consequences, hence desirability, which can only be learned by experimenting with that reform. Similarly, decisions in committees are often analyzed under the assumption that payoff distributions are perfectly known to committee members at the time of their decisions, although this is rarely the case in practice. This difference is important if those decisions are made repeatedly, because one's vote at any given time must take into account the impact of the current decision on everyone's future preferences and votes.

This paper analyzes how experimentation - the fact that an alternative is chosen, despite its comparatively lower immediate expected payoff, in order to learn more about its value - is affected by the nature and amount of individual control over collective decisions. This raises another, dual question: how does the possibility that individual rankings of social alternatives evolve through learning, potentially resulting in majority shifts, affect collective decisions? The analysis is conducted in a two-arm bandit model (settings with multiple risky actions are discussed at the end of the paper). The "safe" alternative yields a constant, homogeneous payoff to all. The "risky" alternative yields payoffs according to some distribution, or *type*, which varies across individuals. At any time, only one of the two action is taken, according to some voting rule. Individuals may learn their type through experimentation with the risky alternative. The paper first analyzes a benchmark setting, in which the type is either "good" or "bad", then extends several results to general structures of preference uncertainty, which accommodate a rich set of applications.

In the benchmark setting, if an individual's type is good, the risky alternative pays him some lump-sums whose arrival times are exponentially distributed. If his type is bad, the risky action pays him nothing. Therefore, an individual knows for sure that his type is good as soon as he receives a lump-sum (he is then a "sure winner"). However, he remains uncertain about his type as long as he has not received anything, for this could be either due to a bad type, or to a late lump-sum arrival for a good type (he is then an "unsure voter"). At each instant, society chooses a single action according to some fixed voting rule. When society experiments with the risky action, this results in a better assessment of individual valuations for that action, hence a better knowledge of individual rankings of alternatives.

A key feature of the analysis is the feedback effect occurring between individual preferences and collective decisions. Indeed, not only does preference uncertainty affect society's choices, but the reverse is also true. This social phenomenon has been described in the context of conservatism by Kuran (1988):

> I believe that a complete model for the study of conservatism would have a circular dynamic structure, with individuals' choices driven by their beliefs and preferences; society's choices generated by its members' choices; and, completing the circle, these members' beliefs and preferences influenced by society's choices. It would thus incorporate three interactive processes: that by which individuals' seek and integrate information to form their beliefs and preferences regarding the alternatives they face; that by which society combines these choices to select policies, institutions, and technologies; and finally, that by which collective outcomes mold individuals' beliefs and preferences. I am suggesting that the aim of theoretical analysis on the subject should be to elucidate these three processes with an eye toward deriving propositions as to when, how, and to what extent individuals and collectivities adapt to changes in environmental factors.

Conservatism resulting from the interaction between preference uncertainty and collective decisions has been studied, both empirically and theoretically, in the context of trade liberalization.[1] Fernandez and Rodrik (1991), in particular, are motivated by the behavior of industry groups who lobby against trade reforms ex ante, but a majority of which benefits from these reforms once they are implemented. They explain this paradox by showing that reforms having a negative expected value ex ante may turn

[1] See Baldwin (1985), Bhagwati (1988) and Fernandez and Rodrik (1991).

out to be beneficial to the majority once implemented. This will be the case, for example, if a reform generates a few "losers," whom it severely hits, while providing small benefits to a majority of "winners". In such a scenario, when the identity of winners and losers is a priori unknown, the reform is initially opposed by all, but eventually gains the support of a majority as the identity of winners and losers gets revealed.

This paper provides a general analysis of the above questions. It shows in particular that the value of experimentation is reduced, compared to a single-decision-maker setting, by one's having to share control over future decisions with other society members. This control-loss may be decomposed into a *loser-trap* effect: an individual has the risky action imposed on him even though it turns out to be bad for him, and a symmetric *winner-frustration* effect: an individual is blocked from enjoying the risky action in the long run, despite his benefiting from it. These effects, each of which occurs with some probability when control is shared, are absent from single-decision-making problems. They reduce the attractiveness of experimentation. The loser trap effect arises even when learning is infinitely fast, so that experimentation entails no time cost: given the possibility of immediately revealing everybody's type, society may prefer to reject this possibility and impose instead the status quo forever, provided the loser-trap effect is strong enough.

Moreover, provided that types are independently distributed, control-loss effects lead to the stark result that, as group size goes to infinity, *everyone votes myopically.* The value of experimentation vanishes and each individual votes for the risky action only if its expected payoff is higher than that of the safe action. Intuitively, this can be explained by the fact that control over future decisions is infinitely diluted and individual power vanishes.

In the benchmark setting, society also experiments *too little compared to the utilitarian policy.* Intuitively, whenever a majority of unsure voters imposes the safe action, it ignores the utility of sure winners. Utilitarian welfare, however, would take sure winners' utility into account, making the risky action more attractive overall than from the sole viewpoint of unsure voters. This result can be reinterpreted in terms of commitment. If individuals are ex ante identical (i.e. behind a veil of ignorance) and able to commit at the outset to some collective policy, they choose the utilitarian policy, since their expected utilities are identical and proportional to utilitarian welfare. Therefore, society experiments less under majority voting than it would if it could commit to a collective policy behind a veil of ignorance. In the symmetric setting where bad news (negative lump-sums) reveals sure losers, and remaining voters are unsure, the opposite bias occurs to some extent: as long as they have the majority, unsure voters push experimentation too far compared to the socially efficient level, as they ignore the welfare of sure losers (this case is more complex however, see Section 6.3). Therefore, for applications that are appropriately modeled by either of these settings, it is possible to identify systematic biases from majority-based experimentation, relative to socially efficient decision making.

Given these results, it is tempting to shift to a normative analysis and ask whether another voting rule would be more beneficial to society than the simple majority rule. For example, if the risky action requires unanimity, an individual cannot has the risky action imposed on him if it turns out to be detrimental for him. Unanimity thus completely gets rid of the loser trap effect. However, this very fact also makes experimentation less attractive. Indeed, any winner is less likely to enjoy the risky action in the long run, for this would require that all other society members also turn out to be winners. Requiring unanimity for the risky action thus exacerbates the winner-frustration effect. Whatever control is gained from being able to veto the risky action is balanced by a control loss for enforcing that same action. Examples indeed show that with the unanimity rule, experimentation may last longer or shorter than under the majority rule.

In contrast to the above results, the paper also shows for the benchmark setting that, even when preferences are independently distributed, so that no learning can occur from the observation of others' payoffs, *good news for any individual is good news for all, and prompts society to experiment more.* This result may seem counter-intuitive, as the occurrence of a new winner brings unsure voters closer to the brink, where risky action is imposed on them forever. However, it also makes unsure voters more likely to enjoy the risky action if they turn out to be winners, an effect that must dominate whenever society experiments in the first place.

As should be expected, the severity of the control-loss effect diminishes if types are positively correlated ex ante, so that interests are more likely to be aligned. Such positive correlation also makes experimentation more attractive as individuals learn from both their and other individuals' payoffs, which reduces the time cost of experimentation. This observation is particularly relevant for large societies composed of groups with high intra-group correlation (see also Section 7).

For several applications envisaged in this paper, such as national security and global warming (Section 2), a more general description of preference uncertainty is required. The paper shows, assuming that types are independently distributed across voters, that collective experimentation is *always* shorter than experimentation in otherwise identical single-person decision making problems, or than what any individual would want if he had dictatorial power over decisions (without assuming type independence). This result holds for all voting rules and for general specifications of the risky action. Moreover, provided that collective decision rules are not "adverse", in the sense that for any individual, the risky action is more likely to be implemented if that individual benefits from it than otherwise, there is always *some* amount of experimentation. In particular, as long as any given individual has some control over decisions and other voters are not encouraged to elect an action whenever it turns out to be detrimental to him, that individual wishes to experiment with the risky action, i.e. to try it even in some circumstances where its immediate expected payoff rate is below the safe action's rate.

The present analysis contributes to the literature on collective conservatism. In contrast to earlier literature, it does not rely on arguments such as exogenous transaction costs or sunk costs or bounded rationality, as surveyed by Kuran (1988). Some results of this paper are closely related to Fernandez and Rodrik (1991), who were the first to interpret the status quo bias as a consequence of preference uncertainty. In their setting however, there is no experimentation. Whatever individuals learn in the first period has no impact on the collective decision of the second period, which is known from the outset, as there is no aggregate uncertainty. Moreover, preferences are immediately fully revealed to all agents, which rules out any dynamic analysis, such as understanding how a positive or negative news shock for one individual affects other voters' willingness to experiment.

The paper is also related to a developing literature on games and experimentation in which conservatism may arise as a consequence of strategic information acquisition, as described by Bolton and Harris (1999), Décamps and Mariotti (2004) and Keller, Rady, and Cripps (2005). Those papers identify a free-riding problem which may result in lower experimentation. Li (2001) provides a theory of conservatism based on a similar argument. In those papers, agents make individual investments to acquire information about the common value of some alternative. In particular, information acquisition amounts to a public-good problem. In contrast, the present analysis considers the reverse setting, in which a single collective action is made at any time, but the value of the action may vary across individuals. This paper therefore also contributes to the literature on experimentation by analyzing voting and experimentation.

More generally, the paper contributes to the experimentation literature in economics, started by Rothschild (1974), Jovanovic (1979), Weitzman (1979), and Roberts and Weitzman (1981). For a recent survey of this literature, see Bergemann and Välimäki (2006). With respect to this literature, the present paper is the first to consider voting and experimentation.

The analysis of the benchmark setting owes conceptual and technical clarity to the use of exponential bandits. Exponential bandits were introduced by Presman and Sonin (1990) and Presman (1990), and used in economics by Malueg and Tsutsui (1997), Bergemann and Hege (1998, 2001), Décamps and Mariotti (2004) and Keller, Rady, and Cripps (2005). In contrast to the last and most closely related paper, where many asymmetric equilibria may occur, the setting considered here has a unique equilibrium which is solved by backward induction on the number of winners observed. This equilibrium is robust to the chosen equilibrium concept, whether it be majority voting equilibrium used by, among others, Roberts (1989), or the weak-dominance solution concept, pervasive in the literature on legislative bargaining.

Section 2 presents a simple example providing intuition for several results of the paper, and describes several applications. Section 3 analyzes the benchmark setting with majority voting and independent types. Section 4 analyzes other decision processes: i) utilitarian policy, ii) commitment, delegation, and delays, and iii) experimentation based on the unanimity rule. Section 5 analyzes the effect of voter heterogeneity and ex ante type correlation on experimentation. Section 6 has two purposes: i) introduce settings with negative news shocks, useful for several applications, and ii) extend several results to general structures of preference uncertainty. Section 7 discusses some assumptions and extensions of the main model, in particular: i) the case of privately observed payoffs, ii) the case of multiple risky actions. Section 8 concludes.

2 Example and Applications

2.1 A Simple Example

Three friends, Ann, Bob, and Chris, go to a restaurant once every week-end. Each week-end, they choose their restaurant using the majority rule. A new restaurant has just opened. Should the friends try it? Do they try it? Suppose the alternative is a restaurant that gives utility 1 to all. For each voter, the new restaurant can be either bad (yielding 0 utility) or good (yielding utility $u > 1$). Suppose that preferences, or "types" are independently distributed across friends (e.g. Ann is no more or less like likely to appreciate the new restaurant if Bob likes it, etc.), with both types having an *ex ante* probability of 1/2.

Immediate Full Type Revelation Suppose that, if they try this new restaurant, all voters immediately learn their type. With probability 1/8, Ann and Bob will like it but Chris won't. In this case, Chris is trapped into always returning to that restaurant, as Ann and Bob have the majority. This situation will be referred to as the "loser trap". Also with probability 1/8, Chris is the only one who turns out to like the restaurant, but is blocked from exploiting this discovery for future dinners by Ann and Bob. This symmetric situation will be referred to as "winner frustration". Overall, there is a probability 1/4 that Chris loses control over the decision process, compared to the situation in which he could choose the restaurant by himself in the future. Depending on u and on how time is discounted, these control-loss effects may be such that Chris and, by symmetry, all voters prefer not to try the new restaurant even though each of them would have preferred to try it if he had full control over future decisions.

Gradual Type Revelation Now suppose that a voter likes the new restaurant if and only if he finds a dish there that he really likes. In this case several visits to the restaurant may be needed to find out one's type. This can lead to situations in which friends experiment with that restaurant until either a majority of them likes it, or a majority of them judges unlikely that they will find anything like there. With this assumption, suppose that, in their first try, only Chris discovers that he likes the new restaurant. What effect does it have on Ann and Bob? Does this incite them to try it more or, on the contrary, prompts them to block new experimentation? Good news for Chris reduces the risk of winner frustration for Ann and Bob, but increases the probability of the loser trap. It turns out that good news for Chris *always* makes Ann and Bob more willing to experiment, as shown in Section 3.

Social Efficiency What would a social planner, wishing to maximize the sum of utilities of the three friends, choose to do? Suppose that u is very close to 1, so that "winners" (those who like the new restaurant) appreciate it only slightly more than the incumbent. Then, the only case in which a social planner would impose the new restaurant in the long run is if all friends turn out to be winners. If there is a loser (i.e. someone who dislikes the new restaurant), the very small utility gain achieved by winners does not compensate the disutility experienced by the loser. However, this policy is incompatible with majority voting, which would result in the two winners imposing the new restaurant despite the much larger magnitude of the loss incurred by the third, losing voter. This difference may result in all friends voting against the new restaurant when their preferences are still unknown, due to the loser-trap effect, while a social planner will try it to see whether all friends like it. In fact, Section 4 shows that, with positive news shocks, majority-based experimentation is always less than the socially efficient policy.

2.2 Applications

The effects described in this paper can arise whenever decisions are made collectively and repeatedly. Although the examination of any particular application is beyond the scope of this paper, the reader may keep in mind the following contexts when thinking about the phenomena analyzed in the following sections.

Reforms with Unknown Losers Reforms constitute a natural domain of application of the above analysis. Even when they benefit a significant fraction of the population, reforms usually harm some individuals or groups. Whenever the identity of these losers is ex ante unknown, the "loser trap effect" becomes a source of conservatism, as this paper shows. Such reforms include the case of trade liberalization studied by Fernandez and Rodrik. Uncertainty regarding the role that each individual will play in a more open society can also justify some lack of popular support for transitions from rigid economic systems to more open ones.

Ex Ante Public Goods Section 6 considers a general specification of preference uncertainty that encompasses a setting, symmetric to the benchmark setting, in which negative lump-sums occur if the action is bad, while the status quo entails a small cost compared to the payoff of the risk risky action if it turns out to be good. This setting captures applications in which a costly effort is required of group members to prevent catastrophes from happening, but the identity of those who may suffer from these catastrophes, if no preventive action is taken, is unknown.

In the context of global warming, for example, the safe action is immediate adoption of drastic policies to cut emissions of greenhouse gases. The risky action is to "explore" the effects of global warming. The safe action has an economic cost. The risky action results in losers, which are the countries most adversely affected by global warming. The "winners" are those countries which suffer least from global warming. Actual consequences of global warming are largely unknown in magnitude and nature. Despite some degree of predictability, the identity of losers and winners is to a large extent also unknown.[2] The effects, whatever they are, are likely to be lasting. Moreover, fighting global warming requires coordination: all significant polluters must simultaneously reduce gas emissions, as emissions emanating from any country affect all other countries. From this viewpoint, the analysis of the paper yields several observations and predictions (see Sections 6.1, 6.3, 4.1, and 4.2). It predicts two possible regimes, depending on parameter values (payoffs, learning speed relative to discount rate, group size, etc.), for gas emissions policies when countries cannot commit to a coordinated long-term policy. In the first regime, individual polluting countries wait to better understand

[2] To take one example, the impact of global warming on several northern European countries will dramatically depend on its influence on thermohaline circulation, which is largely unknown.

the consequences of global warming until, possibly, countries which become actually harmed by its consequences gain enough power to impose drastic emission reductions to all polluters. In the second regime, individual countries, considering the risk of entering the first regime, agree very early on to drastically reduce emissions of greenhouse gases, without learning anything about the actual consequences of global warming. Both regimes are socially inefficient. The first regime pushes experimentation with global warming too far, as polluting countries ignore the woes of affected countries. The second regime does not explore enough the consequences of global warming, as it imposes drastic actions to avoid a phenomenon that remains completely unknown. The paper shows that social efficiency would require a unanimous long-term commitment to a gas emission policy that depends on the observed consequences of global warming. This may entail, for example, all polluting countries agreeing to drastically cut emissions as soon as any one of them is significantly harmed, where this last event may be assessed based on different environmental indicators. The paper also shows that commitment to an observation-dependent policy is very different from commitment to an action, such as an irrevocable imposition of drastic reduction of gas emissions to all polluters. Indeed, commitment to an action reduces efficiency even further than equilibrium policies without commitment, as it adds even more rigidity to the decision process. Furthermore, the paper shows that without policy commitment, coordinated reduction of greenhouse gases i) is implemented earlier than what any country would prefer if it could dictate current and future gas emission policies to all countries, and ii) always involves *some* experimentation, in the sense that for all voting rules, countries always wait to assess, to some extent, the effects of global warming before reducing gas emissions. (See Section 6.)

As another example, some security policies benefit from coordination at an international level (more generally a global, rather than local, level). All countries concerned with the policy must bear the cost of this coordination. Investing in efficient security coordination amounts to taking a "safe" action. The risky action is not to implement such security coordination. The losers are countries who suffer from attacks which might have been prevented or reduced by global coordination. The "winners" are those countries who do not get attacked despite the lack of global security policies. The identity of losers is ex ante unknown, at least to some extent. To fit the dynamic analysis of this paper, it is important that losers are affected in such a way that, after their first attack, they wish, more than unaffected countries, to implement global coordination. In such case, this paper makes similar predictions as in the context of global warming, namely i) policy coordination suffers the inefficiency of the two regimes discussed in the context of global warming, ii) coordination is implemented earlier, others things equal, than what any individual country would prefer if it could dictate to others the degree, present and future, of coordination in security policies, and iii) there is always *some* experimentation, in the sense that for all voting rules, countries always try to assess to some extent the risk of attacks before implementing security coordination.

JOINT INVESTMENT As an illustration of the winner frustration effect, suppose that a firm considers a new, challenging project, which involves several development tasks. Each task has some probability of success. Tasks are complementary: the project can only succeed if all of the tasks are successful. In such a situation, the managers of the individual tasks are exposed to the risk of the project being abandoned due to failures with other tasks of the project, even if they achieve success with their own task. The winner frustration effect dominates this setting, as the gain from the risky action (here, undertaking the project) is large compared to the status quo. The loser trap effect is small here, as losers from the risky action are not significantly harmed. This highlights an important feature of many applications: the tension between a necessary complementarity in actions with a possible misalignment of preferences.

3 Benchmark Setting

As a benchmark, this section embeds the exponential bandit model of Keller, Rady, and Cripps "KRC" (2005) into a setting with majority voting. Time $t \in [0, \infty)$ is continuous and discounted at rate $r > 0$. There is an odd number $N \geq 1$ of individuals who continually decide at the simple majority rule which of two actions to choose. The first action S is "safe" and yields a flow s per unit of time to all individuals. The second action R is "risky" and can be, for each player, either "good" or "bad." The types (good or bad) are independently distributed across the group (the case of correlated types is considered in Section 5). If R is bad for some individual i, it always pays him 0. If R is good for i, it pays him lump-sum payoffs at random times which correspond to the jumping times of a Poisson process with constant intensity λ. The arrival of lump-sums is independent across individuals. The magnitude of these lump sums[3] equals h . If R is good for i, the expected payoff per unit of time is therefore $g = \lambda h$. The assumption $0 < s < g$ rules out the uninteresting case in which either R or S is dominated for all beliefs. Each individual starts with a probability p_0 that R be good for him. This probability is the same for all and is common knowledge. Thereafter, all payoffs are publicly observed, so that everyone shares the same belief about any given individual's type (for privately observed payoffs, see Section 7). In particular, the arrival of the first lump-sum to a given individual i makes him publicly a "sure winner". At any time t, therefore, the group is divided into k "sure winners" for whom R is good with probability one, and $N - k$ "unsure voters," who have the same probability p of having a good type. Unsure voters' probability evolves according to Bayes' rule and obeying the dynamic equation $dp/dt = -\lambda p(1-p)$ if no lump-sum is observed, with p_j jumping to 1 if some voter j receives a lump sum (see KRC, p. 45). Type independence implies that an unsure voter only learns from his payoff stream but not from others'.

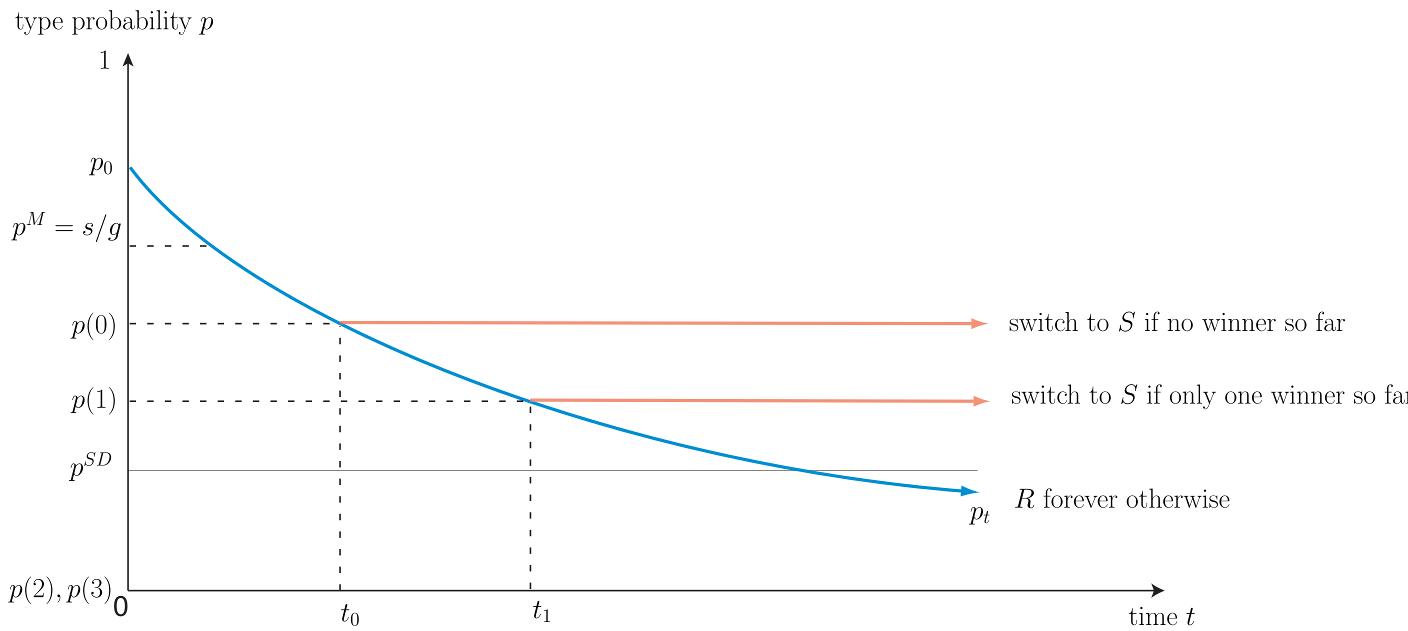
When $N = 1$, the setting reduces to the optimization problem of a single decision maker. The optimal experimentation strategy is Markov with respect to the current belief p, determined by a cut-off p^{SD} such that R is played if and only if $p \geq p^{SD}$. This cut-off equals

$$p^{SD} = \frac{\mu s}{\mu g + (g - s)}, \tag{1}$$

where $\mu = r/\lambda$ (see KRC). Let $p^M = s/g$ denote the myopic cut-off, i.e. the probability below which R yields a lower expected payoff than S. The previous formula implies that $p^{SD} < p^M$. Indeed, experimentation really only takes place for all $p \in [p^{SD}, p^M]$, since the single decision maker then chooses the risky action, despite its lower payoff, in order to learn more about its true value for future decisions. Choosing R in this range is optimal due to the option value of experimentation.

For a group, the first results of this paper show that, with the simple majority rule,

[3] All results hold if these lump sums have random, independently distributed magnitudes with constant mean h. More generally, what matters for decision makers are expected payoff rates of each action and the probability that the risky action be good or bad. See Section 6 for a general specification of payoff distributions and beliefs.



t_0: experimentation end time if no winner is observed before reaching $p(0)$.

t_1: experimentation end time if only one winner is observed before reaching $p(1)$.

$p(2) = 0$: R is elected forever if winners have the majority, no matter what p_t for the remaining unsure voter.

$p^{SD} < p(1)$: a single decision maker always experiments more than a group with a majority of unsure voters.

Figure 1: Dynamics of Collective Experimentation with 3 Voters.

collective decisions are determined by a vector of nonincreasing cut-offs $p(k)_{0 \leq k \leq N}$ such that the risky action is played at time t if and only if $p_t > p(k_t)$, where k_t is the number of sure winners at that time. The dynamics of collective decisions can thus be described as follows. Starting with some (high enough) level p_0 that is common to all, R is elected until the threshold $p(0)$ is reached, at which time two things can happen: either no one received any lump-sum so far, and the safe action is then elected forever. Or at least one winner has been observed by then, in which case experimentation continues until at least another threshold $p(1) < p(0)$, and so on. The dynamics of collective decisions, which is next formally analyzed, is qualitatively represented by Figure 1 for the case of three voters. The intuition for cut-off monotonicity is presented before Theorem 2.

A collective decision rule (or policy) is a stochastic process $C = \{C_t\}_{t \geq 0}$ adapted to the filtration generated by the arrival of voters' lump sums and taking values in the action space $\{R, S\}$. Any collective decision rule determines a value function for each agent i:

$$V_t^{i,C} = E_t \left[\int_t^{\infty} e^{-r(\tau-t)} d\pi^i_{C_\tau}(\tau) \right],$$

where $d\pi^i_S(\tau) = sd\tau$, and $d\pi^i_R(\tau) = hdN^i_\tau$ or 0 depending on whether R is good or bad for i, where $\{N^i\}_{1\le i\le N}$ is a family of independent Poisson processes with intensity λ.

At any time t, within each subgroup of voters (sure winners or unsure voters), all voters have the same value function since their payoffs are identically distributed. Let $w^{k,C}$ and $u^{k,C}$ respectively denote the value functions of sure winners and unsure voters where superscripts indicate the current number k of sure winners and the rule C that is followed. Letting $k_N = (N-1)/2$, winners have the majority if and only if $k > k_N$.

Definition 1 *C is a* Majority Voting Equilibrium *(MVE) if for all t, it satisfies the following conditions:*

- *if $k_t \le k_N$, C solves*

$$\begin{aligned} u_t^{k_t,C} = \sup_\theta E_t \Bigg[& \int_t^\sigma e^{-r(\tau-t)} d\pi^u_{\theta_\tau}(\tau) \\ & + e^{-r(\sigma-t)} \left(\frac{1}{N-k_t} w_\sigma^{k_t+1,C} + \frac{(N-k_t-1)}{N-k_t} u_\sigma^{k_t+1,C} \right) \Bigg], \end{aligned} \tag{2}$$

- *if $k_t > k_N$, C solves*

$$w_t^{k_t,C} = \sup_\theta E_t \left[\int_t^\sigma e^{-r(\tau-t)} d\pi^w_{\theta_\tau}(\tau) + e^{-r(\sigma-t)} w_\sigma^{k_t+1,C} \right], \tag{3}$$

where σ is the first (possibly infinite) time at which a new winner is observed, and θ is any policy.

This definition means that at any time, the subgroup with the majority follows the policy that is optimal for itself, until a change occurs in the composition of the subgroups. When unsure voters have the majority, the conditional probability that any given unsure voter be that new winner is simply $1/(N-k_t)$, since there are $N-k_t$ unsure voters with identical payoff distributions. This explains the last term in (2). This definition extends to a non-Markov setting the standard notion of majority voting equilibrium for dynamic Markov policies (see for example Roberts (1989)). In particular, if one imposes at the outset that the collective decision rule only depend on the state (k, p), the above equations then reduce to the following Hamilton-Jacobi-Bellman (HJB) equations (to be explained in detail shortly):

- If $k \le k_N$, $C = R$ if and only if

$$pg + \lambda p[w^C(k+1,p) - u^C(k,p)] + \lambda p(N-k-1)[u^C(k+1,p) - u^C(k,p)] - \lambda p(1-p)\frac{\partial u^C(k,p)}{\partial p} > s \quad (4)$$

- If $k > k_N$, $C = R$ if and only if

$$g + \lambda p(N-k)[w^C(k+1,p) - w^C(k,p)] - \lambda p(1-p)\frac{\partial w^C(k,p)}{\partial p} > s \quad (5)$$

The equilibrium concept also corresponds, in a dynamic, continuous-time setting, to the usual concept of weak dominance equilibrium: it is the outcome obtained if any member of the majority chooses his optimal action, as if he were pivotal, given that the same will be true of majority members at any future date. The concept rules out trivial Nash equilibria, such as equilibria in which all individuals vote for the same action. It also gets rid of some problems and subtleties specific to continuous games, such as those identified by Simon and Stinchcombe (1989)

The first result of this paper states that there exists a unique[4] majority voting equilibrium, that this equilibrium has the Markov property, and that it is determined by cut-off policies. Existence of Markov equilibria has been widely studied and is common in dynamic stochastic games. However, the fact that any equilibrium of the present dynamic game is Markov, and that there exists indeed a unique such equilibrium, is rare and noteworthy. It owes to the particular structure of the model, which makes possible the use of a backward induction argument on the number of winners, and works despite the infinite horizon, continuous-time nature of the model.

Before formally stating and proving the result, some intuition for it may be helpful (assuming for now the Markov property). At any time t the state of the group can

[4] As usual in the continuous-time stochastic control literature, uniqueness of the optimal policy is understood up to a subset of times of measure 0 on which actions can take any possible values without affecting value functions.

be summarized by k_t and p_t. Each subgroup (sure winners or unsure voters) consists of individuals with perfectly aligned interests. Majority belongs to either sure winners if $k > k_N$ or to unsure voters if $k \leq k_N$. The group with the majority can enforce whichever action it prefers. For example, if sure winners have the majority, it is clearly in their interest to impose R forever. Similarly, if unsure voters have the majority and p is equal (or very close) to 0, so that they are in fact (almost) sure that the action is bad for them, they will impose the status quo S forever, and no further learning occurs. Since an unsure voter can become a winner, but the reverse is false, majority can only shift from unsure voters to winners. Starting with a majority of unsure voters, decisions are dictated by unsure voters' interest until they (possibly) lose the majority. The main question is therefore to determine unsure voters' preferences. These preferences are assessed by the following Hamilton-Jacobi-Bellman (HJB) equation:

$$ru(k,p) = \max\Big\{pg + \lambda p[w(k+1,p) - u(k,p)] + \lambda p(N-k-1)[u(k+1,p) - u(k,p)] - \lambda p(1-p)\frac{\partial u}{\partial p}(k,p), s\Big\}. \quad (6)$$

The first part of the maximand corresponds to action R, the second to action S. The effect of action R can be decomposed into four parts: i) the expected flow rate pg, ii) the jump of the value function if i receives a lump-sum, which occurs at rate λ with probability p: his value function jumps to w and the number of winners increases by 1, iii) the jump of i's value function if another unsure voter receives a lump-sum: i is still an unsure voter, but the number of sure winners increases by 1, and iv) the effect of Bayesian updating on the value function when no lump-sum is observed. Independence of the Poisson processes governing individual payoffs implies that only one lump-sum can be received during any infinitesimal period of time, so that no term involving two or more jumps appears in the HJB equation. In comparison, if S is chosen, learning stops, and i simply receives the flow rate s. Since unsure voters have identical value functions, they unanimously decided to stop experimentation if p becomes too low. They do so when the R part of (6) equals s. At such level p, the smooth pasting condition implies that the derivative term vanishes since the value function is constant, equal to s/r, below that level (see for example Dixit (1993)). This determines the equilibrium policy's cut-offs.

Theorem 1 (Existence and Uniqueness) *There exists a unique MVE. This equilibrium is characterized by cut-offs $p(k)$, $0 \leq k \leq N$, such that R is chosen in state*

(k, p) *if and only if* $p > p(k)$.

Proof. Suppose that $k = N$, i.e. all voters are sure winners. Then, σ is necessarily infinite, so (3) reduces to

$$w_t^{N,C} = \sup_\theta E_t \left[\int_t^\infty e^{-r(\tau - t)} d\pi^w_{\theta_\tau}(\tau) \right].$$

The (essentially) unique solution is $C_\tau = R$ for all τ, since it provides winners at any time with the maximal possible expected payoff g. This gives them the constant value function $w_t^N = g/r$. The value function of unsure voters is also easily computed: if an unsure voter's type is good, which happens with probability p_t, he gets the same expected value as winners, g/r. Otherwise, he gets 0 forever. Therefore, $u_t^N = p_t g/r$. For $k = N - 1$, (3) reduces to

$$w_t^{N-1,C} = \sup_\theta E_t \left[\int_t^\sigma e^{-r(\tau - t)} d\pi^w_{\theta_\tau}(\tau) + e^{-r(\sigma - t)} g/r \right],$$

where I use the fact that $w_t^N = g/r$. Again, the (essentially) unique solution is $C_\tau = R$ for all τ, value functions still equal $w_t^{N-1} = g/r$ and $u_t^{N-1} = p_t g/r$. By the same induction argument, $C_\tau = R$ and $w_t^k = g/r$ and $u_t^k = p_t g/r$ for all $k > k_N$. Now consider the case $k = k_N$, in which unsure voters have the majority, but only one new winner among them is needed for the majority to switch to sure winners. Then (2) reduces to

$$u_t^{k_N,C} = \sup_\theta E_t \left[\int_t^\sigma e^{-r(\tau - t)} d\pi^u_{\theta_\tau}(\tau) + e^{-r(\sigma - t)} \left(\frac{1}{N - k_N} \left(\frac{g}{r} + h \right) + \frac{(N - k_N - 1)}{N - k_N} \frac{p_t g}{r} \right) \right], \tag{7}$$

using the relations $w^{k_N+1} = g/r$ and $u^{k_N+1} = p_t g/r$. The optimization problem (7) is formally identical to the optimization problem of a single decision maker, with known termination values. The solution of such problem is well known (see for example Fleming and Soner (1993)). The control is Markov in p, with any indifference threshold $p(k_N)$ determined by the smooth-pasting condition of the Hamilton-Jacobi-Bellman equation (6), which reduces to

$$pg + p\lambda(g/r - s/r) + p\lambda(N - k_N - 1)(pg/r - s/r) = s, \tag{8}$$

using the relation $u^{k_N+1}(p) = pg/r$. The left-hand side of (8) is increasing in p, equal to 0 if $p = 0$ and higher than $g > s$ if $p = 1$. Therefore, the equation has a unique root, which can be reexpressed as

$$p(k_N) = \frac{\mu s}{\mu g + (g - s) + (N - k_N - 1)(p(k_N)g - s)}. \tag{9}$$

This shows that $C(p, k_N) = R$ if and only if[5] $p > p(k_N)$. If $p \leq p(k_N)$, S is chosen by unsure voters. Since no more learning occurs, p remains constant forever, hence S is played forever. The above strategy entirely determines the value functions $w^C(k_N, p)$ and $u^C(k_N, p)$ of sure winners and unsure voters, which are in fact computable in closed-form by integration of their dynamic equation:

$$w(k_N, p) = \frac{g}{r} - \frac{g - s}{r}\left(\frac{1 - p}{1 - p(k_N)}\right)^{N - k_N}\left(\frac{\Omega(p)}{\Omega(p(k_N))}\right)^{\mu}, \tag{10}$$

and

$$u(k_N, p) = \frac{pg}{r} + \frac{s - p(k_N)g}{r}\left(\frac{1 - p}{1 - p(k_N)}\right)^{N - k_N}\left(\frac{\Omega(p)}{\Omega(p(k_N))}\right)^{\mu} \tag{11}$$

for $p \geq p(k_N)$, where $\Omega(p) = (1 - p)/p$. These functions are easily shown to be increasing in p, with $u^C(k_N, p) \geq pg/r$. Moreover, $u(k_N, p) = w^C(k_N, p) = s/r$ for $p \leq p(k_N)$, since the status quo is imposed forever.

Now suppose that $k = k_N - 1$. Then, any new winner results in the case $k = k_N$ just analyzed. Again, (2) is formally equivalent to the stochastic control problem of a single decision maker. Using again the smooth pasting property in (6), which implies that the derivative of the value function vanishes, any indifference threshold $p(k_N - 1)$, must solve

$$pg + p\lambda(w(k_N, p) - s/r) + p\lambda(N - k_N - 2)(u(k_N, p) - s/r) = s. \tag{12}$$

Since the left-hand side is increasing in p, equal to 0 for $p = 0$ and above s for $p = 1$, the equation has a unique root $p(k_N - 1)$. The choice rule thus defined entirely determines value functions $u(k_N - 1, \cdot)$ and $w(k_N - 1, \cdot)$.

To show that $p(k_N - 1) > p(k_N)$, suppose that the contrary holds. Then, $u(k_N, p(k_N - 1)) = w(k_N, p(k_N - 1)) = u(k_N - 1, p(k_N - 1)) = s/r$, and by the smooth-pasting property, $\frac{\partial u_C}{\partial p}(k_N - 1, p(k_N - 1)) = 0$. Therefore, (12) becomes $p(k_N - 1)g = s$, which contradicts the assumption that $p(k_N - 1) \leq p(k_N) < p^M$. Thus, necessarily, $p(k_N) < p(k_N - 1)$.

[5] As before, this is up to action changes on a time subset of measure 0.

Let us now show that $u(k_N - 1, p)$ is nondecreasing in p. Suppose that $p_t = \tilde{p} > \bar{p}$ and that unsure voters behave as if p_t were equal to $\bar{p}$, meaning that they will stop experimenting after the same amount of time σ_S, unless a new winner is observed σ_W. Until $\sigma = \min\{\sigma_S, \sigma_W\}$, unsure voters receive nothing since R is played and no new winner is observed. The value function of this strategy is thus equal to

$$u(p_t) = E_t \left\{ e^{-r(\sigma - t)} \left[q \left(\frac{1}{N - k_N + 1}(w(k_N, p_\sigma) + h) + \frac{N - k_N}{N - k_N + 1} u(k_N, p_\sigma) \right) + (1 - q)\frac{s}{r} \right] \right\},$$

where $q = Prob[\sigma_W < \sigma_S | p_t]$. We saw that $u(k_N, \cdot)$ and $w(k_N, \cdot)$ are increasing in p. Moreover, these values are above s/r. Indeed, s/r is the value achieved if voters chose the status quo, which is suboptimal by definition of σ_S and given that $p(k_N) < p(k_N - 1)$. Also, p_σ is increasing in p_t given the Bayesian updating dynamics. Finally, σ_W is decreasing in p_t, since a higher p_t makes it more likely that a payoff will be observed.[6] This also implies that q is increasing in p_t by definition of q and the fact that σ_S is independent of p_t by construction. Combining the above implies that $u(\tilde{p}) > u(\bar{p})$. Since unsure voters optimize their value function with respect to σ_S, this yields $u(k_N - 1, \tilde{p}) \geq u(\tilde{p}) > u(\bar{p}) = u(k_N - 1, \bar{p})$, which proves monotonicity of $u(k_N - 1, \cdot)$. $w(k_N - 1, \cdot)$ is also increasing in p_t. Indeed, let $\sigma_1 < \sigma_2$ the arrivals times of lump-sum to the next two new winners. As is easily shown, these stopping times are decreasing in p_t, in the sense of first order stochastic dominance. This, given the fixed experimentation thresholds $p(k_N)$ and $p(k_N - 1)$, implies that the distribution of the (possibly infinite) stopping time σ_S at which experimentation stops increases in p_t in the sense of first-order stochastic dominance. Finally, since

$$w(k_{N-1}, p_t) = E_t \left[\frac{g}{r} \left(1 - e^{-r(\sigma_S - t)} \right) + \frac{s}{r} e^{-r(\sigma_S - t)} \right],$$

this shows that $w(k_{N-1}, \cdot)$ is increasing in p_t. The remaining of the proof proceeds by backward induction on k, where the induction hypothesis is that i) for all $k' > k$, $C(k', p) = R$ if and only if $p > p(k')$, where ii) $p(k')$ is non-increasing for $k' > k$, and iii) the resulting value functions $u(k', \cdot)$ and $w(k', \cdot)$ are non-decreasing in p. The general induction step is then proved exactly as above. ■

The main result of the next theorem is cut-off monotonicity: the larger the number of winners, and the more remaining unsure voters are willing to experiment. This result

[6] Conditional on p_t, σ_W is the mixture of exponential variables with intensity λj, $j \in \{0, \ldots, N - k_N + 1\}$, with mixture weights $\{\rho_j\}$ corresponding to the binomial distribution $B(N - k_N + 1, p_t)$. Monotonicity is in the sense of first-order stochastic dominance.

is perhaps surprising: why would unsure voters want to experiment more when the risk that they lose majority and be imposed R forever increases? The intuition is as follows. Suppose that p is below the myopic cut-off p_M but above $p(k)$ so that with k current winners, unsure voters choose to experiment. By definition of p^M, unsure voters get, in such situation, a lower immediate expected payoff rate with R than with S. Therefore, the only reason why they choose to experiment is that they hope to become winners. Now suppose by contradiction that $p(k+1) > p(k)$, and that p lies in (p_k, p_{k+1}). Then, as soon as a new winner is observed, k jumps to $k+1$, which implies that the status quo is imposed forever, since $p < p_{k+1}$. Therefore, the very reason why unsure voters wanted to experiment, namely the hope of being winners, becomes moot: as soon as one of these unsure voters becomes a winner, he sees the safe action imposed on him forever, which prevents him from actually enjoying any benefit of being a winner.[7] In fact Theorem 3 shows that not only does experimentation increase when a new winner is observed, but the value function of unsure voters also increases, as long as this new winner does not gives majority to sure winners (i.e. as long as $k < k_N$). Another important result contained in the next theorem is that $p(k) > p^{SD}$ for all $k \leq k_N$, which means that a single decision maker would always experiment more than a group whose majority consists of unsure voters. The reason is the control-loss effect mentioned in the introduction: when a single decision maker ignores his type, he still knows that i) if he turns out to be winner, he will be able to enjoy the high-payoff, risky action forever, and ii) if he turns out to be a loser, he can stop experimentation whenever he wants. Neither of these facts are true for unsure voters in a group: even if an unsure voter turns out to be a winner, he is not guaranteed that the risky action will be played forever, since a majority of unsure voters may block it. And if he turns out to be a loser, he may still be imposed the risky action forever if experimentation lasts long enough to reveal that a majority of voters are winners. This twofold control loss prompts unsure voters to experiment less than anyone of them would if he could dictate decisions in the future.

Theorem 2 (Cut-Offs Relations) *Equilibrium cut-offs satisfy the following relations:*

[7] That is, apart from receiving a lump-sum at the time of jump, but the possibility of that gain is already factored in the computation of the immediate expected payoff, which is still less than s for $p < p^M$.

- $p^M > p(0)$.

- $p(k) > p(k+1)$ *for* $k \leq k_N$.

- $p(k_N) \geq p^{SD}$ *with strict inequality if* $N > 1$.

- $p(k) = 0$ *for* $k > k_N$.

Proof. Theorem 1 already shows that $p(k) = 0$ for $k > k_N$. The fact that $p(k_N) \geq p^{SD}$ with strict inequality if $N > 1$ comes from the comparison of (9) and (1). Monotonicity of $p(k)$ is part of the induction in the proof of Theorem 1. There remains to show that $p^M > p(0)$. The indifference condition for $p(0)$ is

$$p(0)g + p(0)\lambda(w(1, p(0)) - s/r) + p(0)\lambda(N-1)(u(1, p(0)) - s/r) = s. \tag{13}$$

Since $p(0) > p(1)$, unsure voters strictly prefer experimentation at $p = p(0)$ when $k = 1$. Therefore, $u(1, p(0)) > s/r$. Since winners always get a higher expected payoff than losers no matter what action is chosen, $w(1, p(0)) \geq u(1, p(0))$. Therefore, the second and third terms on the left-hand side of (13) are positive, which implies that $p(0)g < s$, or equivalently that $p(0) < p^M$. ■

When learning is extremely fast, a single-decision maker is always willing to experiment until he learns his type (almost) perfectly. Mathematically, this result comes from the single-decision maker cut-off equation (1): as the intensity λ goes to infinity, μ goes to 0 and so does the cut-off p^{SD}. However, this result does not extend to the case of collective experimentation. In this case, the time cost of experimentation is only one of two reasons for preferring the status quo. The other reason is the risk of being imposed the risky action forever by the majority while being a loser. If the control-loss effect is strong enough, society may prefer to shun the opportunity of learning everyone's type and make a perfectly informed decision (clearly what a utilitarian planner would choose!), and stay in the dark, i.e. stick forever to the status quo. Keeping other parameter values fixed, this will happen if the total number N of individuals is large enough, as individual power gets more diluted. Mathematically, this immediate revelation does systematically occur if cut-offs stay bounded away from 0 as learning intensity λ goes to infinity. The next result, which provides condition under which this happens, is a direct consequence of (9).

Corollary 1 (Immediate Type Revelation) *If* $N > 2g/s - 1$,

$$\lim_{\lambda \to \infty} p(k_N) = \frac{(N+1)s/g - 2}{N-1} > 0.$$

If $N \le 2g/s - 1$,

$$\lim_{\lambda \to \infty} p(k_N) = 0.$$

Corollary 1 suggests that the total number N of individuals plays an important role on experimentation. Indeed, the next proposition provides a stark result as N gets large. Still assuming type independence, it shows that, at the limit, *individuals behave myopically*, choosing the risky action only if its immediate, expected payoff is larger than that of the safe action. For large groups with independent types, therefore, true experimentation, understood as the election of an action despite a lower immediate payoff in order to learn more about it, completely disappears. To state this result, let $p(k, N)$ denote the experimentation cut-off when there are k winners and N overall individuals.

Proposition 1 (Group Size) $p(k_N, N)$ *is nondecreasing in* N*. Moreover, for all* k*,* $p(k, N) \to p^M$ *as* N *goes to infinity.*

Proof. The first part of the proposition is an immediate consequence of (8). For the second part, (8) also implies that $p(k_N, N) \to s/g = p^M$ as N goes to infinity. To conclude the proof, observe that from Theorem 2, $p(k_N, N) \le p(k, N) \le p^M$ for fixed k and all $N \ge 2k + 1$. Taking the limit as N goes to infinity proves the result. ■

Figure 2 illustrates cut-off policies for different values of N and of the number $\kappa = k_N + 1 - k$ of switches required for winners to gain the majority. In general, cut-offs $p(k, N)$ are *not* monotonic with respect to group size, as can be proved by numerical computation. Such violations may seem counter-intuitive: as N increases, individual power gets more diluted, so shouldn't this reduce the value of experimentation? However, keeping k fixed, increasing N also makes it more likely, for any given unsure voter, that other winners will be observed, for any fixed cut-offs value. Therefore, the addition of new unsure voters reduces the winner frustration effect, for fixed cut-off level. For some parameter values, this may, locally, increase the attractiveness of experimentation. The result of Proposition 1 is therefore not as natural as it may initially appear.

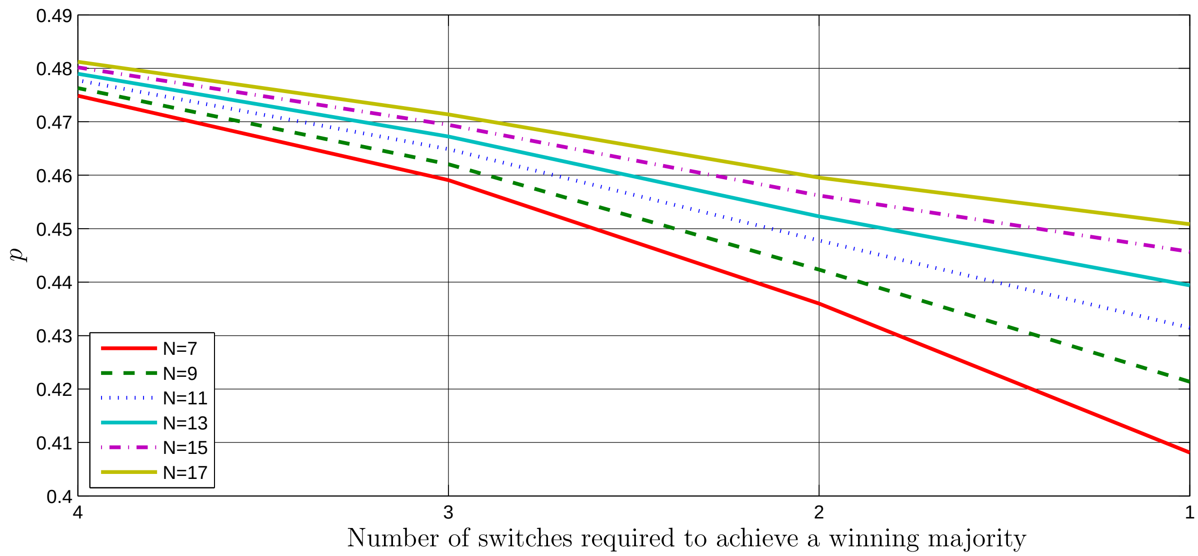


Figure 2: Policy Cut-Offs. $r = 1$, $\lambda = 1$, $s = 1$, $g = 2$.

The previous proposition is also interesting to think about power concentration. Indeed, define an oligarchy as a subset of O (odd) voters such that, at any time, the collective decision is the action chosen by the majority of that subset.

PROPOSITION 2 (OLIGARCHY) *With an oligarchy of O voters, there exists a unique MVE. This MVE is defined by cut-offs such that $C(p_1, \ldots, p_N) = R$ if and only if $p \geq p(k_O, O)$, where k_O is the number of sure winners in the oligarchy.*

Proof. Decisions are entirely determined by the oligarchy. From their viewpoint, the strategic situation is therefore equivalent to a society with only O individuals voting at the majority rule. ■

Combined with Proposition 1, Proposition 2 conveys a sense in which experimentation lasts longer if power is concentrated into fewer hands. In particular, a dictator sets, unsurprisingly, the same experimentation cut-off as the single-decision-maker cut-off p^{SD}.

We have seen in Theorem 2 that society experiments more as the number of winners increases. It is actually possible to prove a stronger result, pertaining to the monotonicity of value functions. Indeed, even *the value function of unsure voters increases* *with the number of sure winners.* For the value function of sure winners, this result is more intuitive: a higher number of sure winners means a higher probability that a winning majority will be achieved. However, to be complete, this argument also requires that experimentation cut-offs decrease in k, which is guaranteed by Theorem 2. Cut-off monotonicity implies that, when a new winner is observed, not only do winners get closer to gaining the majority, but experimentation lasts longer in any case. More surprising is the fact that the occurrence of a new winner is also good news for unsure voters, meaning that their value function jumps upwards, unless this new winner is the decisive voter that gives the majority to winners. The intuition here is that, for $k < k_N$, new winners make experimentation more attractive to unsure voters: if they turn out to be winners, they will be more likely to enforce their preferred action. Of course, it also increases the risk of being imposed that risky action if one turns out to be a loser. However, because unsure voters were already willing to experiment before the new winner is observed, it means that this trade-off was already resolved in favor of experimentation. For $p < p^M$, unsure voters were already in a situation in which the only reason to play the risky action was their hope of being winners. Therefore, the fact that experimentation is facilitated through lower cut-offs and more winners is also good news for them. This argument is contingent on the fact that unsure voters still are in control of the collective decision process, however. When $k = k_N$ and a new winner is observed, remaining unsure voters lose the majority and their value function suddenly drops to the imposed value of the risky action. Finally, as one would expect, value functions are also increasing in p. For unsure voters, this is explained by their higher likelihood of get lump-sums through action R, while their payoff from action S is unchanged. The value function of sure voters is also increasing in p. Indeed, the more likely unsure voters are to be winners, and the longer experimentation will last, in expectation and, hence, the longer sure winners will be able to enjoy high payoffs. These various monotonicity properties are illustrated by Figure 3.

Theorem 3 (Value Function Monotonicity) *The following holds:*

- *u and w are nondecreasing in p,*
- *$w(k,p)$ is nondecreasing in k for all p,*
- *$u(k+1,p) \geq u(k,p)$ for all p, and $k < k_N$,*

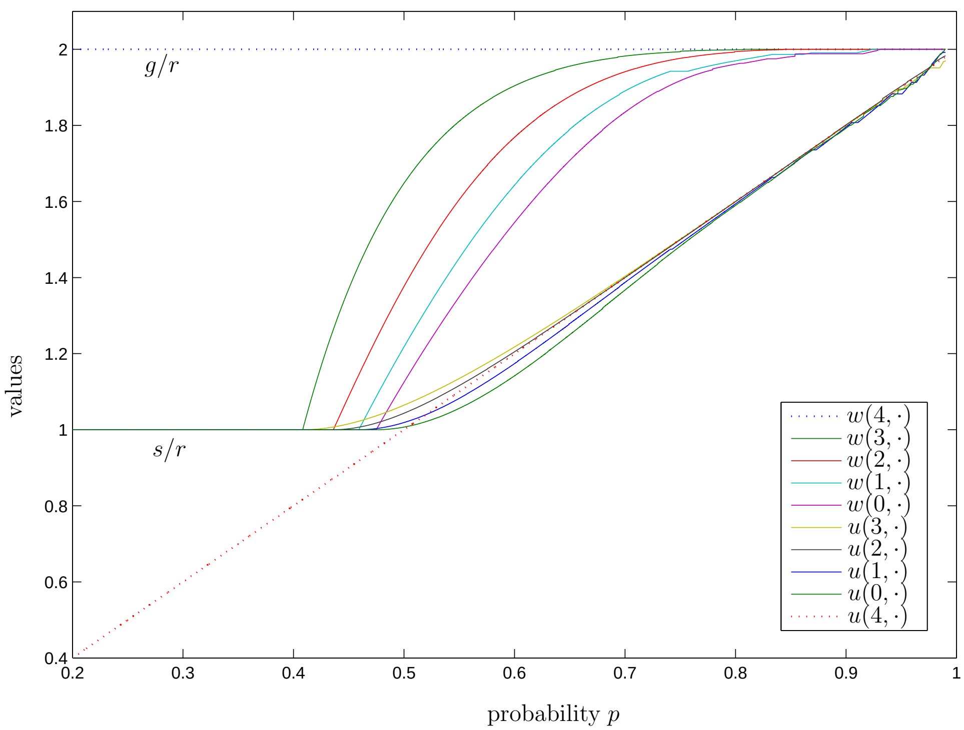


Figure 3: Value Functions. $r = 1$, $\lambda = 1$, $s = 1$, $g = 2$, $N = 7$.

- *$u(k_N+1,p) < u(k_N,p)$ for all p,*
- *$u(k,p) = pg/r$ and $w(k,p) = g/r$ for all p and $k > k_N$.*

Proof. Monotonicity of u and w with respect to p was shown as part of the induction hypothesis in the proof of Theorem 1. If $k > k_N$ R is elected forever since winners have the majority. This determines value functions for this case and yields the last claim. To show monotonicity in k of w for $k \le k_N$, we proceed by induction. Clearly, $g/r = w(k_N+1,p) \ge w(k_N,p)$. Suppose that $w(k,p) \le w(k+1,p)$. We need to show that $w(k-1,p) \le w(k,p)$. Let $\phi(p) = w(k+1,p) - w(k,p) \ge 0$ and $\psi(p) = w(k,p) - w(k-1,p)$. Since $p(k-1) \ge p(k)$, $\psi(p) \ge 0$ for $p \le p(k-1)$. Recall the dynamic equation of w for $p \ge p(k-1)$ and $\tilde{k} \ge k-1$:

$$-rw(\tilde{k},p) + \lambda(N-\tilde{k})p(w(\tilde{k}+1,p) - w(\tilde{k},p)) - \lambda p(1-p)\frac{\partial w}{\partial p}(\tilde{k},p) + g = 0.$$

Taking the difference of the resulting equations for $\tilde{k} = k, k-1$ and rearranging terms yields

$$(r + \lambda p(N-k+1))\psi(p) = \lambda p(N-k)\phi(p) - \lambda p(1-p)\psi'(p).$$

Suppose ϕ is nonnegative by induction hypothesis, the previous equation can be rewritten as $\psi'(p) \le \alpha(p)\psi(p)$ for function α. A direct application of Gronwall's inequality along with $\psi(p(k-1)) \ge 0$ proves that ψ is nonnegative, completing the induction step.

To show monotonicity of u with respect to $k \le k_N$, fix some $k \le k_N$. The dynamic equation of u for $p \ge p(k-1)$ and $\tilde{k} \ge k-1$ is

$$-ru(\tilde{k},p) + \lambda p(w(\tilde{k}+1,p) - u(\tilde{k},p)) + \lambda p(N-\tilde{k}-1)(u(\tilde{k}+1,p) - u(\tilde{k},p)) - \lambda p(1-p)\frac{\partial u}{\partial p}(\tilde{k},p) + pg = 0.$$

Let $\phi(p) = u(k+1,p) - u(k,p)$, $\phi^w(p) = w(k+1,p) - w(k,p)$, and $\psi(p) = u(k,p) - u(k-1,p)$. Taking the difference of the previous equation for $\tilde{k} = k, k-1$ and rearranging terms yields:

$$(r + \lambda p(N-k+1))\psi(p) = \lambda p[\phi^w(p) + (N-k-1)\phi(p)] - \lambda p(1-p)\psi'(p). \qquad (14)$$

We already know that ϕ^w is positive. Therefore, if ϕ were also nonnegative, the argument we just used for w would also show that ψ is nonnegative. In particular, if one can show that $u(k_N, p) \geq u(k_N - 1, p)$, a backward induction will prove the result for all $k \leq k_N$. Combining (10) and (11) implies that, for $k = k_N$,

$$\phi^w(p)+(N-k_N-1)\phi(p) = \frac{g - s - (N - k_N - 1)(s - p(k_N)g)}{r} \left(\frac{1-p}{1-p(k_N)}\right)^{N-k_N} \left(\frac{\Omega(p)}{\Omega(p(k_N))}\right)^{\mu}.$$

Therefore, the left-hand side has the sign of $g - s - (N - k_N - 1)(s - p(k_N)g)$. From the cut-off formula (8), this latter term has the same sign as $s - p(k_N)g$, which is positive. Therefore, we can apply the first term in the right-hand side of 14 is nonnegative for $k = k_N$, which implies that ψ is nonnegative for $k = k_N$. This fills the missing step of the induction, concluding the proof that u is increasing in k for $k \leq k_N$.

To show the last statement, observe that $u(k_N + 1, p) = pg/r$ from Theorem 1, and that $u(k_N, p) > pg/r$, from (11). ■

4 Other Decision Rules

This section investigates how previous results are affected by changes of decision rules, keeping the setting otherwise identical. The optimal experimentation policy of a social planner maximizing utilitarian welfare is first derived and shown to last longer than majority-based experimentation. The impact of commitment, delegation, and delays in decision rules is then considered. In particular, the utilitarian policy is shown to be equivalent to the policy to which all voters would want to commit at the outset if placed behind a veil of ignorance. The distinction between committing to a policy versus committing to an action has dramatic consequences on policy efficiency. Finally, majority-based experimentation is then compared to the unanimity-based experimentation, where it is shown that no voting rule dominates the other in terms of efficiency. In brief, when unanimity is required for the risky action, the loser trap effect disappears, but the winner frustration effect is reinforced, which may reduce experimentation further than the majority rule under circumstances illustrated in this section.

4.1 Utilitarian Criterion

Theorem 4 *Under the utilitarian criterion, the optimal policy is determined by cut-offs $q(k)$ such that $C(k,p) = R$ if and only if $p \geq q(k)$. These cut-offs are non-increasing in k, with $q(k) = 0$ if*

$$k \geq \bar{k} = \frac{s}{g}N.$$

Proof. The proof is similar to that of Theorem 1, proceeding by backward induction on the number k of winners. For $k \geq \bar{k}$, the utilitarian optimum is to choose R forever even if $p = 0$, since sure winners' gains from R outweigh the aggregate gain from S even if all unsure voters get nothing from R. This fact can be expressed as $q(k) = 0$ for $k \geq \bar{k}$. The resulting welfare is $W(k,p) = k\frac{g}{r} + (N-k)\frac{pg}{r}$. Consider next $k = \bar{k} - 1$. Let $w^C(k,p)$ and $u^C(k,p)$ denote the value functions of sure winners and unsure voters if policy C is used, given that R is played forever if a new winner is observed, and let $W^C(k,p) = kw^C(k,p) + (N-k)u^C(k,p)$, denote utilitarian welfare under policy C. Then, the utilitarian criterion C must solve

$$W_t^{k_t,C} = \sup_{\theta} E_t \left[\int_t^{\sigma} e^{-r(\tau - t} \sum_i d\pi^i_{\theta_\tau}(\tau) + e^{-r(\sigma - t)} W_\sigma^{k_t+1,C} \right],$$

where σ is the first (possibly infinite) time at which a new winner is observed, and where $W_\sigma^{k_t+1,C} = W(\bar{k}, p_\sigma)$, the welfare that was computed earlier for $k = \bar{k}$. This is a standard control problem, whose solution is Markov. The indifference boundary must satisfy the smooth pasting condition

$$kg + (N-k)pg + (N-k)\lambda p \left[\frac{kg + (N-k)pg}{r} - \frac{Ns}{r} \right] = Ns,$$

which has a unique root $q(k)$, since the left-hand side is increasing in p, greater than Ns if $p = 1$ and less than Ns for $p = 0$, by definition of $\bar{k}$. Therefore, $C(k,p) = R$ if and only if $p \geq q(k)$. This entirely determines $w(k,\cdot)$, $u(k,\cdot)$ and $W(k,\cdot)$, which are easily shown to be increasing in p. The remaining of the proof proceeds by backward induction on k as in Theorem 1, where the induction hypothesis is that i) for all $k' > k$, $C(k',p) = R$ if and only if $p > q(k')$, where ii) $q(k')$ is non-increasing for $k' > k$, and iii) resulting value functions $w(k',\cdot)$, $u(k',\cdot)$, and $W(k',\cdot)$ are non-decreasing in p. ∎

The next result states that majority experimentation is inefficiently short compared to the utilitarian optimum.

Theorem 5 (Majoritarian vs. Utilitarian Rules) *$q(k) \le p(k)$ for all $k \le k_N$.*

Proof. The utilitarian cut-off $q(k)$ solves

$$(k/N)g + (1 - k/N)pg + (N - k)\lambda p \left[\frac{W(k+1,p)}{N} - s/r \right] = s, \tag{15}$$

while the majoritarian cut-off $p(k)$ solves

$$pg + (N - k)\lambda p \left[\frac{\bar{w}(k+1,p)}{N-k} + \frac{N-k-1}{N-k} \bar{u}(k+1,p) - s/r \right] = s \tag{16}$$

where $\bar{w}$ and $\bar{u}$ are the value functions obtained under the majoritarian rule. Optimality of the utilitarian policy implies that for all k, p, $\frac{W(k,p)}{N} \ge (k/N)\bar{w}(k,p) + (1 - k/N)\bar{u}(k,p)$. Since $\bar{w} > \bar{u}$, this also implies that $\frac{W(k+1,p)}{N} > 1/(N-k)\bar{w}(k+1,p) + (1 - 1/(N-k))\bar{u}(k+1,p)$, and subsequently that the left-hand side of (15) is higher than that of (16) for given p. Therefore, the root of the first equation must be lower than that of the second. ■

4.2 Commitment, Delegation, and Delays

If voters are initially homogeneous and can commit to a policy at the outset, they will choose a policy that maximizes their expected payoffs which are identical and the sum of their expected payoffs. This latter maximization is identical to the utilitarian policy above. This shows the following result.

Theorem 6 (Commitment) *If voters can commit to a policy at time 0, they choose the cut-off policy determined by cut-offs $\{q(k)\}_{0 \le k \le N}$.*

Theorem 6 suggests that social efficiency can be partially restored if voters can commit to a *policy* to some extent. However, such choice should not be confused with commitment to an *action*. Indeed, a policy prescribes which action should be taken depending on past observations. It adapts collective decisions to circumstances. Commitment to an action, in contrast, is harmful because too rigid. The intuition appears best in the extreme case where voters must commit to a once-and-for-all action at the outset. Then, their preference is to choose the risky action if and only if it is above the myopic cut-off. This extreme case of action commitment thus entirely annihilates experimentation. This result can be reinterpreted as follows: if votes take place at a low time frequency, this reduces everyone's individual control over collective decisions even further than in the benchmark case, which causes even less experimentation and thus even more inefficiency. As another consequence, suppose that voters can temporarily transfer the decision process to a delegate who makes decisions based on a mixture of electoral and welfare concerns. Such delegation can improve efficiency to the extent that welfare enters the delegate's objective, as he is able to adapt to incoming information. In contrast, commitment to an action (either safe or risky) reduces willingness to experiment and increases inefficiency.

4.3 Unanimity Rule

Suppose now that R can be enforced only if everyone votes in its favor.

PROPOSITION 3 *Under the unanimity rule, there exists a unique MVE. This MVE is defined by cut-offs $\chi(k)$ which are decreasing in k, and such that $\chi(N-1) = \chi^{SD}$.*

Proof. The first part is proved similarly to that of Theorem 1. For the last part, observe that if $k = N - 1$, the remaining unsure voter has full control over collective decisions. His optimal policy is therefore the same as that of a single decision maker.■

In general, $\chi(k)$ can be smaller or greater (even for $k \leq k_N$) than the majoritarian cut-off $p(k)$. Here are examples illustrating both possibilities.

EXAMPLE 1 *For N large, $p(k_N) \sim p^M$, independently of λ, but $\chi(k_N) \to 0$ as $\lambda \to \infty$ for N fixed.*

Proof. From (8), if N is large, necessarily $p(k_N)$ is close to p^M, *independently of* λ (to see this, divide (8) by $N\lambda$). With the unanimity rule however, fixing N large but finite, and letting λ go to infinity, the experimentation cost for unsure voters goes to 0, as they learn almost immediately their type, and they lose no power. Specifically,

$u(k_N, p) \to p^N g/r + (1-p^N)s/r$ as λ goes to ∞, which is strictly greater than s/r for all $p > 0$. Hence $\chi(k, N) \to 0$ as $\lambda \to \infty$. ■

EXAMPLE 2 *Suppose that $N = 3$ and $s \ll g$. Then $\chi(1) > p(1)$.*

Proof. Equation (8) implies that

$$p(1) = \frac{\mu s}{\mu g + (g-s) - (s-pg)} \sim \frac{\mu s}{(\mu+1)g} \tag{17}$$

if $g \gg s$. In particular, $p(1)$ is very close to zero if $g \gg s$. On the other hand, indifference of unsure voters with unanimity the rule, and $k = 1$ obtains if p satisfies

$$pg + \lambda p[w(2,p) - s/r] + \lambda p[v^{SD}(p) - s/r] = s, \tag{18}$$

where $w(2,p)$ is the value of a sure winner under unanimity rule if there are two sure winners (and $N = 3$), and $v^{SD}(p)$ is the value function of a single-decision maker. As can be easily checked, $v^{SD}(p) \le pg/r + (1-p)s/r$, while $w(2,p) \le pg/r + (1-p)s/r$. This and (18) imply that $\chi(1)$ must satisfy the inequality

$$pg + 2\lambda p^2(g/r - s/r) \ge s,$$

or

$$p \ge \frac{\mu s}{\mu g + 2p(g-s)} \sim s/g \tag{19}$$

if $g \gg s$. Comparing (17) and (19) shows that $\chi(1) > p(1)$. ■

5 Correlation and Heterogeneity

This section considers the case of two voters, 1 and 2, who share a common belief about the initial joint distribution of their types, although this distribution need not be symmetric any more, and allows for correlation between the voter types. Let θ^i denote Voter i's type, and let $p^{t_1 t_2} = Pr[(\theta^1, \theta^2) = (t^1, t^2)$, where $t^i \in \{g, b\}$ represent the possible types (good or bad) of each voter. Also let $p^i = Prob[\theta^i = g]$ for $i \in \{1, 2\}$, and $\alpha = p^{gg}/(p^1 p^2)$. α is a measure of the correlation between voter types. The usual correlation measure and α have a one-to-one relationship for any given p^1 and p^2. If

$\alpha = 1$, types are uncorrelated. In general, α takes values in $\mathbb{R}_+$, although not all values of $\mathbb{R}_+$ are achievable for given p^1, p^2. For example, $p^1 = 1$ implies that $\alpha = 1$, since in that case Voter 1's type is deterministic hence uncorrelated with Voter 2's type. Let Δ denote the set of (p^2, α) that are achievable as elementary probabilities vary over the four-dimensional simplex. The following proposition is a simple exercise of Bayesian updating, whose proof is easy and omitted.

Proposition 4 (State Dynamics) *Beliefs are governed by the following dynamics equations. When no lump-sum is observed,*

- $\frac{dp^{gg}}{dt} = -\lambda p^{gg}(2 - p^1 - p^2)$
- $\frac{dp^{bb}}{dt} = \lambda p^{bb}(p^1 + p^2)$
- $\frac{dp^{gb}}{dt} = -\lambda p^{gb}(1 - p^1 - p^2)$, $\frac{dp^{bg}}{dt} = -\lambda p^{bg}(1 - p^1 - p^2)$
- $\frac{d\alpha}{dt} = -\lambda\alpha(1 - \alpha)(p^1 + p^2)$

When Voter 1 receives a lump-sum,

- $p_+^{bb} = 0$, $p_+^{gb} = \frac{p^{gb}}{p^1}$, $p_+^{bg} = 0$, $p_+^{bb} = \frac{p^{bb}}{p^1}$
- $\alpha_+ = 1$, $p_+^1 = 1$, $p_+^2 = \alpha p^2$

where the subscript '+' denotes values immediately after the lump-sum is observed, and its absence denotes values immediately before the lump-sum. Symmetric formulas if instead Voter 2 receives a lump sum.

With two voters, let us replace the majority rule by assuming that unanimity is required to play R. Since voters may now be heterogeneous (i.e. $p^1 \neq p^2$ even if none of them is a sure winner, the concept of equilibrium must be modified. In the spirit of Section 3, and given the unanimity rule, it is natural to assume that the voter who is the less likely of being a winner (i.e. voter i if $p^i \leq p^j$), is in control: if that voter wants to play the risky action, so should the player with a higher expected type. This notion is also consistent with elimination of weakly dominated strategies, because the pivotal voter is always the voter who wishes to stop experimentation. For simplicity, let us therefore define a unanimity equilibrium (UE) as follows: at any time t, if $p^i \le p^j$, then j votes for R whenever i does.

Theorem 7 *There exists a unique UE. This equilibrium determined by a cut-off function $\delta : \Delta \to [0,1]$ such that $C(p^1, p^2, \alpha) = R$ if and only if $p^1 > \delta(p^2, \alpha)$ whenever $p^1 \le p^2$, with the reverse relation if $p^1 > p^2$.*

Proof. First suppose that $p^2 = 1$. Then voter 1 has full control over the collective decision. He therefore imposes his optimal policy, which is that of a single decision maker. This defines $\delta(1,1) = p^{SD}$. This also fully determines the value functions of both voters in that case. Let $p \mapsto w(p)$ denote the value function of voter 2, where p is voter 1's probability of being a winner is p, and v^{SD} is the value function of a single decision maker, which is also voter 1's value function in this case. More generally suppose that at time 0, $p_0^1 \le p_0^2$. It follows from Proposition 4 that $p_t^1 \le p_t^2$ for all t preceding the first arrival of a lump-sum. In particular, this implies that 1 has full control of the collective decision (under unanimity) over that period. Therefore, he chooses a policy θ that solves

$$u_t = \sup_\theta E\left[\int_t^\sigma e^{-r(\tau-t)} d\pi^1_{\theta_\tau}(\tau) + e^{-r(\sigma-t)}\left(qw(p^2_{\sigma_+}) + (1-q)v(p^1_{\sigma_+})\right)\right],$$

where, letting σ_i denote the (possibly infinite) time at which i receives his first lump sum, $\sigma = \min\{\sigma_1, \sigma_2\}$ and $q = Prob[\sigma_1 < \sigma_2]$. This is a standard control problem, whose solution is known to be Markov. Voter 1 is indifferent between R and S at probability level p, if p solves the equation

$$pg + \lambda p[w(\alpha p^2) - s/r] + \lambda p^2[v^{SD}(\alpha p) - s/r] = s. \tag{20}$$

The left-hand side is increasing in p, equal to 0 for $p = 0$ and greater than $g > s$ if $p = 1$. Therefore, it has a unique root $\delta(p_2, \alpha)$. This shows that $C(p^1, p^2, \alpha) = R$ if and only if $p^1 > \delta(p^2, \alpha)$. The case $p^1 > p^2$ obtains by symmetry. ■

The next theorem shows that a voter's incentive to experiment increases both with the other voter's probability of being a winner and with voters' type correlation. The latter result is intuitive: if types are more positively correlated, the control-loss effect weakens: if a voter turns out to be a winner, he is less likely to be blocked by the other

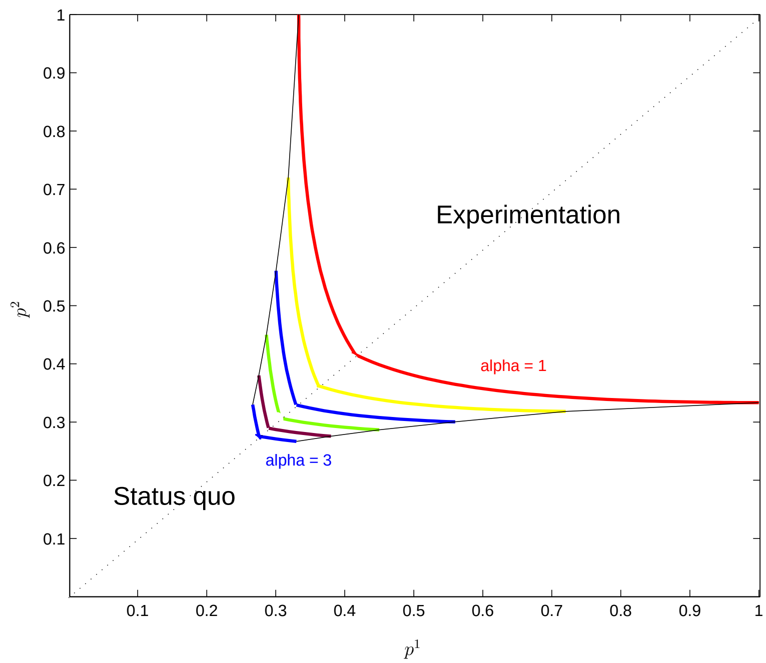


Figure 4: Experimentation Boundary δ as a function of α. $r = 1$, $\lambda = 1$, $s = 1$, $g = 2$.

voter. The first result is also intuitive, given the unanimity rule, a higher probability by the other voter of being a winner makes it less likely that one's decision be blocked. In the extreme case in which, say, Voter 2 is a sure winner (i.e. $p^2 = 1$), Voter 1 has full control over collective decisions, and can behave in effect as a single-decision maker. In addition, positive correlation increases the speed of learning, since voters learn from both their own and the other voter's payoff observation. In the extreme case of perfect type correlation, the setting is equivalent to a single decision maker setting in which the learning intensity is the double of individual arrival rates. Faster learning reduces the time-cost of experimentation, hence makes it more attractive and reduces cut-offs. These effects yield the next result, which is illustrated by Figure 4 representing the experimentation domain as a function of the correlation measure α.

THEOREM 8 *δ is decreasing in both components.*

Proof. The left-hand side of (20), is increasing in p, p^2 and α. Therefore, keeping α fixed, the root $\delta(p^2, \alpha)$ must be decreasing in p^2, and similarly keeping p^2 fixed, $\delta(p^2, \alpha)$

must be decreasing in α. ■

6 Negative Shocks and General Bandits

The purpose of this section is twofold. First, it considers the opposite of the benchmark setting, where news shocks are negative. This setting is useful for several applications, as illustrated by Section 2. Second, it generalizes two results of the previous sections: i) a group always performs less experimentation than any of its member would if he could dictate future decisions, and ii) there is always *some* experimentation, provided that decision rules are not *adverse*, in the sense that the preferences of any voter are, at any time, positively taken into account by society. Finally, another setting, allowing for both positive and negative shocks is considered, in which population at any time is divided between sure winners, sure losers, and unsure voters.

6.1 Negative News Shocks

Several applications mentioned in Section 2 require a setting in which news events amount to bad news. To accommodate such applications, suppose that the risky arm pays a positive constant rate if it is good and, in addition, also pays some negative lump sums according to some Poisson process if it is bad. One may assume without loss of generality that the safe rate is zero, since all payoffs can be translated by the same constant without affecting voters' decision problem.

In that case, the state is composed of k sure losers and the probability p that the arm be good for other voters. Moreover, p increases in time, reflecting the fact that, in this setting, no news is good news for unsure voters.

It can be shown that the policy is also determined by cut-offs $\rho(k)$ such that unsure voters impose the risky action if and only if $p \geq \rho(k)$. Since unsure voters' belief p_t increases over time, the risky action can only be stopped, if used at all, when enough sure losers are observed, either because those obtain the majority and impose the safe action, or because the cut-offs $\rho(k_t)$ get high enough and jump over p_t.

This setting is also useful to visualize the general theorems to follow. Moreover, each theorem has a corollary for the setting with negative news shocks, as summarized in Section 6.3.

6.2 General Bandits

Suppose that, for any given individual, the risky arm has a payoff distribution or "type" θ lying in some finite set Θ. At any time, that individual's belief about his type is summarized by a probability distribution or "state" $\gamma \in \Gamma$, where $\Gamma = \Delta(\Theta)$ is the set of all probability distributions[8] over Θ. The safe arm still pays a constant rate s. For a single decision maker, the *Gittins index* of the risky arm is the map $G : \Gamma \to \mathbb{R}$ such that, given state γ, $G(\gamma)$ is the smallest value of s for which the single decision maker prefers the safe action over experimentation. Mathematically, $G(\gamma)$ solves

$$G(\gamma) = \inf \left\{ s : s/r = \sup_{\sigma} E\left[\int_0^{\infty} e^{-rt} d\pi_{\sigma_t}(t) | \gamma, s \right] \right\},$$

where σ is any policy, and the expectation is conditional on the current state γ and on the rate s of the safe action.[9]

Now consider the case of N decision makers. Still assuming publicly observed payoffs, let $\{\mathcal{F}_t\}_{t\geq 0}$ denote the filtration generated by all voters' payoffs. At any time, the state, known to all, is denoted γ. If types are independent, then $\gamma = \left(\gamma^1, \ldots, \gamma^N\right) \in \Gamma^N$. In general, γ may contain more information, such as type correlation (see Section 5). A policy is a process adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and taking values in $\{S, R\}$.

For any rate s, policy C, and voter i, necessarily

$$\sup_{\sigma} E\left[\int_0^{\infty} e^{-rt} d\pi^i_{\sigma_t}(t) | \gamma, s \right] \geq E\left[\int_0^{\infty} e^{-rt} d\pi^i_{C_t}(t) | \gamma, s \right]. \tag{21}$$

The inequality obtains because C is an element of the policy set over which the maximization is taken.[10] We may define a policy-dependent generalization of the Gittins

[8] In the benchmark model, the type θ is either "good" or "bad" and the state γ is the probability p that the type be good.

[9] Although not explicitly stated, the results of this section naturally apply to discrete-time settings.

[10] In general, C depends on all voters' types and need not be anonymous.

index as

$$G_C^i(\gamma) = \inf\left\{s : s/r = E\left[\int_0^\infty e^{-rt} d\pi_{C_t}^i(t)|\gamma, s\right]\right\}.$$

Inequality (21) implies that $G_D^i(\gamma) \geq G_C^i(\gamma)$ for all i, γ, and C, where $G_D^i(\gamma)$ is i's Gittins index if he has dictatorial power over all decisions.

The definition of voting equilibria is extended as follows. The ν-supermajority rule is the map $\upsilon : \{S, R\}^N \to \{S, R\}$ such that $\upsilon = S$ if and only if the number of votes for S is greater than or equal to some constant ν.

Definition 2 (Voting Equilibria) *C is a voting equilibrium for voting rule υ if for all γ, $C = S$ if and only if the number of voters i such that $G_C^i(\gamma) \leq s$ is greater than or equal to ν.*

The following result shows that collective experimentation stops earlier than individual experimentation in some qualified sense.

Theorem 9 *Suppose that C is a voting equilibrium for voting rule υ. Then, $C = S$ whenever $|\{i : G_D^i(\gamma) \leq s\}| \geq \nu$.*

The proof is an immediate consequence of the general inequality $G_D^i(\gamma) \geq G_C^i(\gamma)$ for all i and, C and γ.

Theorem 9 states that experimentation stops whenever there are at least ν voters, each of which would prefer to stop if he had dictatorial control over the policy. When types are independent, then $G_D^i(\gamma) = G(\gamma^i)$, where $G(\gamma^i)$ is the Gittins index of the single decision maker problem with state γ^i. In that case, i's optimal policy is independent of other individuals' types. Therefore, Theorem 9 implies that, in the case of independent types, experimentation stops whenever at least ν voters would want to stop if placed in a single decision maker setting, given their individual state. This distinction is important: if types are positively correlated, collective experimentation can last longer than in a single-decision maker setting, as positive type correlation increases learning speed and thus reduces the time cost of experimentation, as shown in Section 5. In contrast, collective experimentation is always less, even with positive correlation, than what any voter would like if he could dictate all decisions. In this case, not only does he avoid the loser trap and winner frustration effects, but he can also exploit other voters' payoffs to learn about his type, which he cannot do in a single decision maker setting.

Theorem 9 is illustrated for the case of negative and mixed shocks in Sections 6.1 and 6.4.

The next result requires the following definitions. For any probability distribution γ^i over the type space, let $g(\gamma^i) = E[d\pi^i_R/dt|\gamma^i]$. $g(\gamma^i)$ is i' immediate expected payoff rate with action R given type distribution γ^i. For any individual type θ^i, let, slightly abusing notation, $g(\theta^i) = g(\delta_{\theta^i})$, where δ_{θ^i} is the Dirac distribution concentrated on type θ^i, denote i's true immediate expected payoff rate with action R when his actual type is θ^i.

Definition 3 *i is a winner (resp. a loser) if*

$$g(\theta^i) > (\leq) s.$$

This generalizes the definitions of winners and losers of Section 3. Individual i is a winner if, given his actual type θ^i, the risky action is better than the safe one, and a loser in the opposite case. Therefore, if i perfectly knows his type, he prefers R if he is a winner and S if he is a loser. The set of types can thus be partitioned into "good" types and "bad" types, such that a type is good if and only if a voter with this type is a winner.

Definition 4 *A policy C is adverse for Voter i if the set*

$$\{t : Pr[C_t = R|\theta^i \text{ good}] < Pr[C_t = R|\theta^i \text{ bad}]\}$$

has positive Lebesgue measure.

A policy is adverse for Voter i if the risky action is more likely to be chosen if his type is bad than if his type is good, at least for some nonzero time set. This may be the case, for example, if i's type is negatively correlated with a decisive majority of voters. In this case, these voters will block the risky action whenever it benefits i, and impose it whenever it hurts him. If types are independently or positively distributed, one expects policies not to be adverse to any voter since, all else equal, a voter's influence over collective decisions results in decisions that favor him. Non-adversity is not as benign an assumption as it may appear. For example, in a setting with both positive and negative news shocks (Section 6.4), unsure voters may, upon observing a sure loser, want to push experimentation further as the risk of the loser trap is reduced, which adversely affects this loser. However, it holds in important particular cases. First, it holds when unanimity is required for either of the actions. Second, it holds in the negative news shock setting, as will be verified in Section 6.3, on which several of applications are based. Finally, it also holds in the mixed-shocks setting of Section 6.4, as long as learning is fast enough or, equivalently, voters are patient enough.

THEOREM 10 *Suppose that C is a voting equilibrium for voting rule v. Then, $G^i_C(\gamma) \geq g(\gamma^i)$ for all i for which C is non-adverse.*

Proof. For any safe rate s and policy C, Voter i's expected payoff with policy C is

$$V^i_C = E\left[\int_0^\infty e^{-rt} d\pi^i_{C_t}(t)\right] = \int_0^\infty e^{-rt} E[d\pi^i_{C_t}(t)], \tag{22}$$

where expectations are conditioned on γ.

$$E[d\pi^i_{C_t}(t)] = Pr[C_t = S]sdt + Pr[C_t = R]E[d\pi^i_{C_t}(t)|C_t = R]$$

Therefore, if $E[d\pi^i_{C_t}(t)|C_t = R] > sdt$ for all t, then $V^i_C > s/r$, implying that $G^i_C(\gamma) > s$. Suppose that $s < g(\gamma^i)$. Then, by definition of $g(\cdot)$ and by the fact that the probability of each type is a martingale, $E[d\pi^i_R(t)] = g(\gamma^i)dt > sdt$. Moreover, C's non-adversity with respect to i implies that

$$E[d\pi^i_{C_t}(t)|C_t = R] \geq E[d\pi^i_R(t)],$$

as will be shown shortly. This inequality shows that $G^i_C(\gamma) > s$ for all $s < g(\gamma^i)$, which proves the theorem. To show the inequality, observe that, by Bayes' rule, C is non-adverse for i if and only if $Pr[\theta^i \text{ good}|C_t = R] \geq Pr[\theta^i \text{ good}|C_t = S]$. Moreover,

$$\begin{aligned} E[d\pi^i_{C_t}(t)|C_t = R] = Pr[\theta^i \text{ good}|C_t = R]E[d\pi^i_{C_t}(t)|C_t = R, \theta^i \text{ good}] \\ + Pr[\theta^i \text{ bad}|C_t = R]E[d\pi^i_{C_t}(t)|C_t = R, \theta^i \text{ bad}]. \end{aligned} \tag{23}$$

Combining these results yields the inequality. ■

6.3 Properties of the Negative-Shock Setting

Theorem 9 implies the following result for the negative-news setting.

COROLLARY 2 *Suppose that types are independent. Then, $\rho(k) \geq \rho^{SD}$ for all $k < N/2$, where ρ^{SD} is the single-decision-maker cut-off.*

Moreover, the equilibrium policy resulting from the majority rule is non-adverse to any voter. Indeed, suppose that the risky action is elected by unsure voters at some time t. Then, the (possibly infinite) time at which the risky action is abandoned necessarily decreases with the number of sure losers, since as time passes, the belief of unsure voters gets better and better (see Section 6.1). Therefore, for any unsure voter i, $Pr[\theta^i \text{ good}|C_t = R] \geq Pr[\theta^i \text{ bad}|C_t = R]$, which shows non-adversity of the equilibrium policy. Theorem 10 then implies that all cut-offs lie below the myopic cut-off, hence that unsure voters always experiment to some extent.

Section 4 showed that majority-based experimentation is inefficiently short compared to the utilitarian policy, in the benchmark setting where news events amount to positive news. With negative news shocks, the reverse intuition holds to some extent: provided unsure voters have the majority, they may be willing to experiment for a wider set of states than a utilitarian social planner who takes into account sure losers. However, the value of the risky action is higher for society as a whole when the utilitarian policy is followed than with the majority-voting equilibrium policy, by social optimality of the utilitarian policy. This effect makes experimentation more valuable for the social planner, and may offset the first effect. For example, suppose that no loser has been observed yet. Then, unsure voters may require a higher probability of success to start experimentation than a social planner would, because the latter can always exploit information to improve social welfare, whereas unsure voters face loser trap and winner frustration effects. As the number of sure losers increases, the first effect starts to dominate, with the social planner stopping experimentation if too many losers are observed. This will be case if i) some losers have been observed, who face large negative lump sums, but ii) unsure voters have a very low probability to be losers. In such circumstances, a social planner will not wish to push experimentation further, whereas unsure voters, ignoring the plight of revealed sure losers, will continue with the risky action.

In the negative-shock setting, cut-offs need not be monotonic in the number of losers. Such violations can be observed numerically or in combination with analytical results omitted here. It is interesting to understand why such violation occurs in the negative-shocks setting but not in the positive-shocks setting. The apparition of a new loser has mixed effects for unsure voters: it reduces the risk of being imposed the risky action, but it also reduces the value of experimentation. The latter effect is similar to the positive-news setting. The main difference is that the risk of being imposed the risky action is more severe in the negative-news setting. Intuitively, unsure voters' control over collective decisions is better in the positive-news case than in the negative-news one. In the positive-news case, the only event that can happen to an unsure voter is to become a sure winner. If this does not happen, unsure voters can impose the status quo at any time (or until they lose the majority, but this only occur if $k = k_N$). In contrast, with negative news, the worst that can happen to an unsure voter is to receive a negative lump-sum and suddenly join the minority of sure losers, which have no control of collective decisions. Thus, negative news is compounded by a sudden control loss. This explains why the "insurance" effect resulting from the apparition of a new loser can, paradoxically, encourage experimentation. Furthermore, if nothing happens in the negative-news case, p simply increases which is enough to push experimentation forward. In contrast, in the positive-news setting, the apparition of news winners is *necessary* for experimentation to continue, for otherwise, p decreases until it causes experimentation to stop.[11]

6.4 Mixed Shocks

Now suppose that the benchmark setting is modified as follows: if the risky arm is good, it pays positive lump sums according to the jumping times of some Poisson process with intensity λ_g, and if it is bad, it pays negative lump sums according to the jumping times

[11] From a technical viewpoint, another distinctive feature of the negative-news settings is that the smooth-pasting property does not hold any more. Indeed, as time elapses, p moves *away* from its threshold $p(k)$, so the value function need not be smooth at that cut-off. Instead, cut-offs are determined by direct comparison of value functions with and without starting experimentation.

of a Poisson process with intensity λ_b. Without loss of generality, also suppose that the safe rate is zero (action payoffs can be translated so as to achieve this condition, without affecting preferences). In this case, there are three categories of voters: sure winners (who received a positive lump sum), sure losers (who received a negative lump sum), and unsure voter. Starting with homogeneous beliefs and independently distributed types, the state at any time can be summarized by three numbers: the number k_W of sure winners, the number k_L of sure losers, and unsure voters' individual probability p that the risky action be good for any one of them. Proceeding as in Section 3, one may show that the unique majority voting equilibrium policy, starting with N odd voters, is determined by cut-offs $p(k_W, k_L)$, that equals 0 if $k_W > N/2$, that equals 1 if $k_L > N/2$, and that lies in $(0, 1)$ in the remaining case, for which unsure voters are pivotal. Theorem 9 implies that $p^D \leq p(k_W, k_L)$, where p^D is the cut-off an unsure voter would use if he had full control over future decisions or, in a single decision maker setting, provided types are independent. This inequality holds for all supermajority rules. Moreover, if the risky action requires the unanimity rule, then Theorem 10 implies that $p(k_W, k_L) \leq p^M$, where p^M is the myopic cut-off. That is, unsure voters always wish to experiment to some degree given the unanimity rule. With the majority rule, Theorem 10 does not apply directly, because non-adversity need not hold.[12] However, as learning becomes faster, an argument similar in spirit to non-adversity also implies that unsure voters always experiment to some extent.

Indeed, suppose that either λ_b or λ_g becomes infinite, so that unsure voters immediately learn their type if they elect the risky action. Let $g > 0$ and $b < 0$ the expected payoff rates of the risky arm for sure winners and sure losers respectively.[13] Then, the expected value of the risky action for any unsure voter i, unsure voters are ex ante pivotal (i.e. $\max\{k_W, k_L\} < N/2$), is

$$\begin{aligned} V^i = {} & Pr[k_W^+ > N/2] \left(g/r Pr[\theta^i \text{ good}|k_W^+ > N/2] + b/r Pr[\theta^i \text{ bad}|k_W^+ > N/2]\right) \\ & + Pr[k_L^+ > N/2] \left(g/r Pr[\theta^i \text{ good}|k_L^+ > N/2] + b/r Pr[\theta^i \text{ bad}|k_L^+ > N/2]\right), \quad (24) \end{aligned}$$

where k_W^+ (resp. k_L^+) denotes the number of winners after all types are revealed. Clearly,

[12] Whether it holds in this case is an open question.

[13] These payoff rates can have any magnitude, since they are the product of jump intensities and lump-sum magnitudes.

$Pr[\theta^i \text{ good}|k_W^+ > N/2] > Pr[\theta^i \text{ good}] > Pr[\theta^i \text{ good}|k_L^+ > N/2]$. This implies that

$$V^i > g/rPr[\theta^i \text{ good}] + b/rPr[\theta^i \text{ bad}] = pg/r + (1-p)b/r,$$

which is nothing by the myopic payoff. Therefore, Voter i is willing to experiment at least until p drops below the myopic cut-off p^M, defined by $p^M g/r + (1-p^M)b/r = 0$ (since $s = 0$).

Since only the ratios r/λ_b and r/λ_g of the discount rate and learning intensities matter for the analysis, the result can be reinterpreted as follows: if voters are patient enough, the majority voting equilibrium always entails some experimentation.

7 Extensions and Discussion

Privately Observed Payoffs

Previous sections assumed that all payoffs were publicly observable. What happens if payoffs are privately observed? The following analysis shows that, perhaps surprisingly, this need not affect experimentation.

First consider the case of two individuals with independent types voting at the unanimity rule. For a winner, playing the risky action is always optimal as it maximizes immediate payoff and prompts the other voter to experiment longer, as indicated by Theorem 8. However, why wouldn't an unsure voter want to wrongfully pretend that he is a winner? Indeed, this would prompt the other voter to experiment more, still by Theorem 8, giving more time to the former voter to check whether he is a winner, while leaving him the possibility to stop at any time.

However, such manipulation has no value. Indeed, suppose that Voter 2 mistakenly believes that Voter 1 is a winner. When would Voter 1 want to stop pretending that he is a winner and impose the status quo? In such scenario, Voter 2 thinks he has full control over the decision process, hence experiments up to the single-decision maker threshold. Therefore, Voter 1 can choose any level of experimentation up to the single decision maker threshold at which Voter 2 will stop if he turns out be a loser. Voter 1 may wish to stop earlier however. His optimal cut-off p is determined by the indifference equation

$$pg + \lambda p[w(p) - s/r] + \lambda p[v^{SD}(p) - s/r] = s,$$

where $w(p)$ the level to which his value function jumps if he receives a lump-sum, and v^{SD} is his new value function (i.e. that of a single decision maker) if Voter 2 receives a lump-sum (see Section 5). This indifference equation is identical to the case of complete information: i) the payoff flow (pg) is the same, ii) the winner value $w(p)$ is the same, since a lump-sum makes Voter 1 an actual winner and iii) a lump-sum to Voter 2 prompts that voter to choose the risky action forever, independently of what Voter 1 had pretended to be. This suggests that truthful type revelation is optimal.

With N voters, the intuition is the same. When unsure voters have the majority and payoffs are publicly observed, they already impose the experimentation level that is optimal to each of them, since their interests are perfectly aligned. Therefore, assuming that other voters are truthful, an unsure voter cannot benefit from manipulating the level of experimentation. Mathematically, his indifference equation is unchanged. This again suggests that truthful type revelation is optimal.

There remains to discuss how individuals communicate their types. An obvious way, if the group is small enough, is to use cheap talk. If the group is large, the following protocol is natural. Suppose that experimentation starts with a common type probability p for all voters. Voters are willing to experiment until the first threshold $p(0)$, even if no one has received any lump sum by then. When $p(0)$ is reached, suppose that voters who are still unsure at that point vote for S, while sure winners vote for R. Upon observing the number of votes for R, unsure voters can deduce whether to continue experimentation or not. If no one has voted for R, the status quo is imposed forever. If $k > 0$ voters have chosen R, then experimentation resumes (i.e. everyone votes for R) until p reaches the cut-off $p(k)$, at which point unsure voters vote for the status quo while sure winners, whose number k' is greater than or equal to k, choose R. If $k' > k$, experimentation continues at least up to $p(k')$. If $k' = k$, experimentation stops at $p(k)$. Also assume that off the equilibrium path if one voter chooses the safe action when he should be experimenting, he is believed to be an unsure voter, and vice versa. The above protocol is not information efficient in the sense that voters only know the true state of the world when p reaches some particular cut-offs. Indeed, other less natural protocols would be more efficient, for example if a new winner reveals his type change by voting for S for an infinitesimal amount of time. However, even the less informationally efficient but more natural protocol suffices to implement the full-information policy, as stated in the next theorem. What matters is that the natural protocol is sufficiently informative to exactly implement the policy of Section 3.

THEOREM 11 *The protocol defined above is an equilibrium of the dynamic voting game with privately observed payoffs.*

Proof. For sure winners, it is clearly optimal to follow the protocol, as it requires them to vote R forever. Indeed, such strategy maximizes their immediate payoff as well as the length of experimentation, due to the cut-off monotonicity established in Theorem 2.

With the protocol, the benchmark policy is exactly implemented. Therefore, unsure voters value function, given k and p is the same as in Section 3. However, under the above protocol, unsure voters (and, less importantly, other voters) only know the true state k when particular cut-offs are reached. Let l denote the last such public release of information. For $p > p(l)$, unsure voters only know that the number $\tilde{k}$ of voters who have received lump sums so far is greater than or equal to l. The symbol tilde is added to indicate that k is random from unsure voters' viewpoint. The first part of the proof is to verify that under the protocol, unsure voters wish to experiment for $p > p(l)$. This is indeed the case if their value function is greater than s/r, the value they get with the safe action. This value function, when the protocol is followed by all, equals $U(l,p) = E[u(\tilde{k},p)|l]$. However, crucially, unsure voters only matter if they have the majority. Conditioning on unsure voters being pivotal, the support of $\tilde{k}$ lies in $K(l) = \{l,\ldots,k_N\}$. By Theorem 3, $u(\bar{k},p) \geq u(l,p)$ for $\bar{k} \in K(l)$. Therefore, $U^{piv}(l,p) \geq u(l,p) > s/r$ for $p > p(l)$, where the superscript piv indicates the conditioning on the event $\tilde{k} \leq k_N$. Therefore, it is optimal for unsure voters to choose the risky action whenever indicated by the protocol. Similarly, if, upon reaching $p(l)$, it turns out that $k = l$, i.e. no new winner has been observed since the last release of public information, then it is optimal for unsure voters to stop. Indeed, if they follow the protocol, their value function is identical to the benchmark case, because the protocol policy is exactly the same as in that case. Therefore, their indifference equation is identical to the benchmark case, yielding the same cut-off $p(l)$.

There remains to show that an unsure voter cannot benefit from misrepresenting his type. This is done by backward induction on k. If $k > k_N$, sure voters impose the risky action forever, so that unsure voter's value function is identical to the benchmark case, and he cannot manipulate the protocol. Now suppose that $k = k_N$. Pretending to be a sure winner prompts unsure voters to continue experimentation beyond the full-information level. From above, however, it is suboptimal from that voter's viewpoint, given his value function and his value function if other sure winners were observed: his HJB equation is identical to the benchmark case by induction hypothesis. Therefore, the unsure voter cannot benefit from deviating if $k = k_N$, resulting in the benchmark policy being implemented in that case, which yields the value function $u(k_N, p)$. By backward induction, suppose that the unsure voter does not benefit from deviating for all k strictly greater than $\bar{k}$ hence that his value function is the same as in the benchmark case for such k. Then, his HJB equation at $p = p(\bar{k})$ and knowing $\bar{k}$ is the same as in the benchmark case, so he cannot benefit from deviations. The benchmark policy is thus also implemented in that case, which yields him a value function $u(\bar{k}, p)$ for all p, concluding the induction step. ■

Factions and Heterogeneous Voting Weights. The analysis of this paper extends naturally to settings in which some voters weigh more in the decisions than others. Given the results of this paper, it is natural to expect that voters with more decision weight will be more inclined to experiment longer. For example, suppose that there are four voters with Voter 1 weighing twice as much as other voters and decisions being made at the simple majority rule. Suppose that Voter 4 is the only sure winner so far. Then, Voter 1 can impose experimentation to the level that he desires since, by siding with Voter 4, he creates a majority for the risky action. Therefore, as long as no new winner is observed, Voter 1 can push experimentation up to the single decision maker threshold. However, he will stop earlier if, say, Voter 2 becomes a winner. Indeed, Voter 1 should then fear the possibility that Voter 3 receives a lump-sum, resulting in a winning coalition that imposes upon Voter 1. Contrary to the benchmark setting, thus, experimentation can be interrupted by the occurrence of a new winner. (In this case, though, experimentation still lasts longer than if Voter 1 were split into two independent voters.)

Risk aversion. Although not mentioned so far, the analysis of this paper does not require that voters be risk neutral. Indeed, voters could have any von Neumann-Morgenstern utility function, where lump sums actually correspond to "lump utils", or certainty equivalents thereof if the magnitude of these lump utils is random.

Side payments. In another paper, work in progress, I show how social efficiency, according to utilitarian welfare, can be "spontaneously" restored in the two-arm bandit setting of this paper, if side payments are allowed.

Coordination Breakdown with Multiple Risky Actions

In settings with multiple, correlated risky actions, even the slightest risk of modification in the preference ranking of some voter can result in the group choosing the most conservative action. The fear of such ranking modification may result in coordination breakdown among group members who would otherwise agree to impose some more lucrative action. The following setting provides a stark illustration of the control-loss effect.

Consider a group three individuals (1,2, and 3) voting at the majority rule. First suppose that there are two actions, R and S. S pays a constant rate $s = 1$ to all players. R pays a certain rate of 2 to Voters 1 and 2, but has an uncertain payoff distribution for Voter 3: with probability p it gives off lump-sums to Voter 3, whose corresponding expected payoff is $g = 0.1$. In this simple configuration, it is easy to see that there is a unique equilibrium, in which Voters 1 and 2 impose R forever. This equilibrium holds independently of p, and goes against Voter 3's interest: even if $p = 1$, Voter 3 prefers S, which gives him a higher payoff than 0.1.

Now suppose that there is another risky action, X, such that i) X is good for Voters 2 and 3 if and only if R is good for Voter 3, ii) X surely pays -9 to Voter 1, and has expected payoffs $g_X^2 = 2.1$ and $g_X^3 = 1.1$ respectively to Voters 2 and 3 if it is good. Perfect correlation implies that any lump-sum observed by Voters 2 or 3 with action X or by Voter 3 with action R causes the common probability p to jump to 1.

Since there are three actions, let us assume that if no action receives at least two votes, then the status quo S is imposed. The impact on equilibrium of preference uncertainty appears through the next propositions.

Proposition 5 *If $p = 0$ or $p = 1$, there is a unique MVE: if $p = 0$ R is played forever, if $p = 1$, X is played forever.*

Proof. If $p = 0$ there is no learning. Individual preferences are ordered as follows: $R \succ_1 S \succ_1 X$, $R \succ_2 S \succ_2 X$, and $S \succ_3 X \sim_3 R$. In particular, R is the unique Condorcet winner hence the unique MVE. Similarly if $p = 1$, there is no learning: individual preferences are ordered as follows: $R \succ_1 S \succ_1 X$ and $X \succ_2 R \succ_2 S$, and $X \succ_3 S \sim_3 R$. X is the Condorcet winner hence the unique MVE. ■

For the next result, suppose that $\mu = r/\lambda = 1$.

Proposition 6 *If $p \in (0.1, 0.8)$, there is a unique MVE. In this MVE, S is played forever.*

Proof. Voter 1 is indifferent between R and S if p solves the equation

$$2 + \lambda p \left[-\frac{9}{r} - \frac{1}{r} \right] = 1,$$

which yields $\underline{p} = 0.1$. Thus if p is greater than $\underline{p}$, Voter 1 prefers the low payoff of S rather than risking that Voters 2 and 3 discover that X is good for them and imposing it forever, from Proposition 5. Voter 3's indifference equation relative to actions X and S is

$$p1.1 + \lambda p[1.1/r - 1/r] = 1,$$

which yields $\bar{p} = \mu/(0.1 + 1, 1\mu) = 0.8$ if $\mu = 1$. Therefore, if $p \in (.1, .8)$, S is the preferred choice for 1 and 3, hence chosen forever. ■

To illustrate the content of these propositions, consider the following variation of the restaurant example of Section 2. Three friends, Chris, Ian, and Paul go the restaurant once every week-end, and choose each time their restaurant according to the majority rule. They start with the following preferences. Chris likes Chinese cuisine above anything else. Ian likes Indian cuisine but not Chinese one. Paul does not know Asian cuisine, and is thus uncertain about his preferences. There are three restaurants in town: i) a gourmet Chinese restaurant, clearly the best choice for those like Chinese cuisine, ii) a Singaporean restaurant whose menu contains both Chinese and Indian dishes, and iii) a plain restaurant who has a well known common, relatively low value to all friends.

For some parameter values of everyone's tastes, the following paradox may occur: Ian and Paul vote for the plain restaurant. However, if the gourmet Chinese restaurant closes down, Chris and Ian vote for the Singaporean restaurant. Why did Ian change his vote? In both cases, Ian prefers the Singaporean restaurant. However, if he agrees with Chris to go there, he runs the risk that Paul discovers that likes Chinese cuisine, resulting in Chris and Paul to impose the gourmet Chinese restaurant in the future. If this risk is high enough, Ian prefers to vote for the plain restaurant, which Paul also prefers if his expected value for Asian cuisines is low enough.

8 Conclusion

The analysis of this paper has shown that, in a dynamic setting, collective decisions tend to be too conservative compared to what any individual would choose if he had full control over future decisions, other things equal. Moreover, when news shocks amount to good news, experimentation is also inefficiently short compared to the utilitarian outcome, with a partial equivalent for the negative-shock setting. These phenomena stem from a twofold control loss: the risk of being imposed the riskier actions when those turn out to be detrimental for oneself ("loser trap" effect), and the opposite risk of not being able to enjoy those riskier actions when one turns out to be a winner, if those are blocked by a conservative majority ("winner frustration" effect). For large groups, these effects can entirely annihilate the value of experimentation, causing individuals to vote myopically, but are reduced if actual preferences types are positively correlated. The analysis also shows how commitment to an observation-dependent policy and commitment to an action have very different implications for efficiency.

References

Baldwin, R.E. (1985) *The Political Economy of U.S. Import Policy*, Cambridge, MA. MIT Press.

Bergemann, D., Hege, U. (1998) "Venture Capital Financing, Moral Hazard and Learning," *Journal of Banking and Finance*, Vol. 22, pp. 703–735.

Bergemann, D., Hege, U. (2001) "The Financing of Innovation: Learning and Stopping," *RAND Journal of Economics*, Vol. 36, No. 4, pp. 719–752.

Bergemann, D., Välimäki, J. (2006) "Bandit Problems," Cowles Foundation Discussion Paper No. 1551.

Bhagwati, J. (1988) *Protectionism*, Cambridge, MA. MIT Press.

Bolton, P., Harris, C. (1999) "Strategic Experimentation," *Econometrica*, Vol. 67, No. 2, pp. 349–374.

Décamps, J.-P., Mariotti, T. (2004) "Investment Timing and Externalities," *Journal of Economic Theory*, Vol. 118, pp. 80–102.

Dixit, A. K. (1993) *The Art of Smooth Pasting,* in Fundamentals in Pure and Applied Economics, Vol. 55. Harwood Academic Publishers.

Fernandez, R., Rodrik, D. (1991) "Resistance to Reform: Status Quo Bias in the Presence of Individual- Specific Uncertainty," *American Economic Review*, Vol. 81, No. 5., pp. 1146–1155.

Fleming, W.H., Soner, H.M. (1993) *Controlled Markov Processes and Viscosity Solutions*, Springer, New York.

Jovanovic, B. (1979) "Job Search and the Theory of Turnover," *Journal of Political Economy*, Vol. 87, pp. 972–990.

Keller, G., Rady, S., and Cripps, M. (2005) "Strategic Experimentation with Exponential Bandits," *Econometrica*, Vol. 73, No. 1, pp. 39–68.

Kuran, T. (1988) "The Tenacious Past: Theories of Personal and Collective Conservatism," *Journal of Economic Behavior and Organization*, Vol. 10, pp. 143–171.

LI, H. (2001) "A Theory of Conservatism," *Journal of Political Economy*, Vol. 109, No. 3, pp. 617–636.

MALUEG, D.A., TSUTSUI, S.O. (1997) "Dynamic R&D Competition with Learning," *RAND Journal of Economics*, Vol. 28, pp. 751-772.

PRESMAN, E.L. (1990) "Poisson Version of the Two-Armed Bandit Problem with Discounting," *Theory of Probability and its Applications*, Vol. 35, pp. 307-317.

PRESMAN, E.L., SONIN, I.M. (1990) *Sequential Control with Incomplete Information*, Academic Press, New York.

ROBERTS, K. (1989) "The Theory of Union Behavior: Labour Hoarding and Endogenous Hysteresis," Mimeo, London School of Economics.

ROBERTS, K., WEITZMAN, M. (1981) "Funding Criteria for Research, Development and Exploration of Projects," *Econometrica*, Vol. 49, pp. 1261–1288.

RODRIK, D. (1993) "The Positive Economics of Policy Reform," *American Economic Review*, Vol. 83, No. 2, pp. 356–361.

ROTHSCHILD, M. (1974) "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, Vol. 9, pp. 185–202.

SIMON, L., STINCHCOMBE, M. (1989) "Extensite Form Games in Continuous Time: Pure Strategies," *Econometrica*, Vol. 57, pp. 1171–1214.

WEITZMAN, M. (1979) "Optimal Search for the Best Alternative," *Econometrica*, Vol. 47, pp. 641–654.