

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Behavioral Biases and Group Decision

Permalink

<https://escholarship.org/uc/item/1jg4v6xg>

Author

Towbin, Peter L

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

BEHAVIORAL BIASES AND GROUP DECISION

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Peter L. Towbin

March 2018

The Dissertation of Peter L. Towbin
is approved:

Professor Daniel Friedman, Chair

Professor Nirvikar Singh

Professor David Draper

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Peter L. Towbin

2018

Table of Contents

List of Figures	iv
List of Tables	v
Abstract	vi
Acknowledgments	viii
1 Introduction	1
2 Culture, knowledge and group induction	5
3 The social induction theory of probability weighting	30
4 Social Choice in the Wild	61
5 Conclusion	87
Bibliography	88

List of Figures

2.1	Simulations: Zero confirmation bias, high belief variance.	21
2.2	Simulations: Higher confirmation bias reduces belief variance.	24
2.3	Greater confirmation bias (.4) reduces erroneous voting.	25
2.4	Simulations w/o noise: Confirmation bias induces group belief polarization. 27	
3.1	Kahneman & Tversky (1992) Probability Weighting Function, W^+ for gains, and W^- for losses.	35
3.2	Kerr & Tindale (2011) models of group performance	36
3.3	Majoritarian shift from individual to group belief.	38
3.4	Probability weighting inverts majoritarian polarization.	43
3.5	Population analysis reveals non-random heterogeneity	48
3.6	A motor lottery uses the subject's own error distribution to construct gambles	50
3.7	Classical descriptive lottery weighting functions on the left, and equivalent motor lotteries of the same subjects on the right. Dark curve is median response.	52
3.8	Relatively little overlap is seen between regions correlated with probability weighting on the classical versus motor tasks.	53
3.9	dlPFC responds strongly in the first half of the experiment, in response to descriptive probabilities, then less in the second half after learning from experience has occurred.	56
3.10	Quite a few regions in the brain have been linked to integration of prior probability information. From Hansen, Hillenbrand and Ungerleider (2012)	59
4.1	Desal Preferences: Left: All door-to-door. Right: The ones who attended	73
4.2	Desal preference shift: small group (left) and big group only (right). . .	74
4.3	Left: Desal preference density of those who went to both small and large. Right: 1)Pre-small group, 2) Pre-small who made it through big group as well. 3) Post-small group, 4) Post small and big for those who did both.	75
4.4	Desal Preferences: Left: All door-to-door. Right: The ones who attended	83

List of Tables

4.1	Number of people and level of participation.	69
4.2	Description of candidates given in the survey.	70
4.3	(Door to door survey) Preferences regressed: desalination versus candidates.	77
4.4	(Pre small group) Preference for desalination regressed against candidate preferences.	77
4.5	(Post small group) Preference for desalination regressed against candidate preferences.	78
4.6	Regression after a reflective follow-up discussion, after small and large group events.	78
4.7	Condorcet results from pre survey for each of the eight small groups.	85
4.8	Condorcet results from post survey for each of the eight small groups.	86

Abstract

Behavioral biases and group decision

by Peter L. Towbin

Human behavioral biases are a topic of great importance within the field of economics, but they also are of significance for a broad range of human endeavors. Within economics we most often encounter these biases in puzzling patterns of individual behavior, as in insurance purchases, affinity for gambling, or stock portfolio choices. But, behavioral biases are also thought to play an important role in many critical aggregate phenomena, from market swings to partisan political deadlock. This thesis presents a model for how several of these biases may have arisen together during evolution and suggests that individual biases are intricately linked to the dynamics of group decision making: 1) confirmation bias acts as a stabilizing influence when individuals provide input to group decisions, holding their input steady in the face of rival motives or distracting information; 2) group decisions tend to accentuate belief polarization; and 3) probability weighting, when seen through this evolutionary lens, simply undoes the polarization that occurred when the shared belief was acquired. The final chapter presents an experimental approach to overcoming behavioral biases in the context of deliberative democracy. Small group dialog often triggers biases that lead to belief polarization, but when properly structured, participants coalesce around causal knowledge of the problem domain, rather than entrenched conceptions of the solution.

Acknowledgments

The winding road that led to this thesis has been at times profound and exhilarating, and at other times exhausting. A number of professors provided inspiration along the way and made it possible to stay the course and complete the journey. I was very fortunate to have three of the most inspirational professors on my committee, and I'd like to express my deep gratitude for their insight as well as patience.

Nirvikar Singh first introduced me to the remarkable puzzle of social choice theory, which played an important role in my thesis and has remained an area of fascination. David Draper shared his deep knowledge of Bayesian statistics, decision theory and rigorous scientific thinking. And, my deepest appreciation goes to Dan Friedman who has overseen the voyage, persistently yet gently nudging it forward. Thank you for your many insights across a remarkable array of topics, and always steering me back to practical questions and achievable goals.

Chapter 1

Introduction

Popular conceptions of economics hew closer to classical economic theory than to its more technical, modern incarnation. How, for example, does the human "propensity to truck, barter and exchange one thing for another", as Adam Smith put it[74], lead to the balancing of production with consumption, and ultimately to the wealth of nations? But, Smith also noted that this propensity is probably "the necessary consequence of the faculties of reason and speech". Pinning down exactly what those faculties of reason are, and rebuilding economic theory upon these first principles, is at the heart of the modern economic paradigm. The workhorse and lynchpin of this paradigm has been rational choice theory and the concept of expected utility. For this reason, fault lines within the rationality assumptions of human behavior have been both controversial and, as evidence accumulated, the subject of intense research interest. We now recognize that humans have a number of systematic biases that diverge from traditional concepts of rationality, yet we are still lacking a well accepted explanation for these

biases.

The initial motivation for this thesis is most directly expressed in the final chapter. The fault line explored there is not so much a failure of individual rationality, but the inability to extend the rational choice paradigm to the actions of groups. The formalization of individual choice opened up an array of mathematical and statistical tools to help humans make decisions when preferences are clear but the complexity of evaluating the options is beyond innate human computational ability. The hope was that a parallel theory of group choice would, in a similar manner, provide tools and guidelines to assist in optimizing the even more difficult task of collective choice, from small group decisions to the operations of democratic governance. But, this branch of economic research, social choice theory, did not reach a very happy ending. Kenneth Arrow proved that the logical structure of rational choice theory cannot be directly scaled into a normative theory of group choice [1], and this made it much more difficult to provide theoretical insights and practical decision tools for group decision making.

On the other hand, since the heyday of social choice theory some decades ago, the rational choice model has come under increasing scrutiny. From both casual observation and introspection expected utility theory appeared self evident to many, but careful experimentation in the lab and aggregate puzzles in the market demonstrated disparities between theory and behavior. If people were not “constrained” to act according to the logic of rational choice, could that be an opening for a resolution of Arrow’s result? Initially, the challenge to group action only looked more grim. Biased information assimilation sometimes leads individuals to grow apart when exposed to the

same information, in contrast to what we would expect from a typical Bayesian model. And, something about group dynamics pushed groups that mostly agreed on an issue to move to a more extreme opinion, rather than converging to the average belief. This was dubbed “the law of group polarization” [78].

This tendency towards polarization, so persistent that some would designate it a law of human group behavior, raised a question, which at first seemed so improbable as to be just a curiosity. Could this bias in group behavior somehow explain one or more of the biases in individual behavior? One of those individual biases, equally puzzling on its own, seemed a potential match: individuals do not treat all probabilities equally. Relatively extreme probabilities, those near but not quite at zero or one, are mysteriously scaled by people in lab experiments to be less extreme. No matter how sure the subjects are of the actual probabilities, they still scale the probability to a less extreme value before using it to make a choice. From an evolutionary perspective humans arose in a highly social environment, mutually dependent on their group-mates, and group decisions, for survival. Perhaps the evolutionary dynamics of group living led to the group polarization tendency, and the unconscious weighting of probabilities seen in the lab is a vestige of an adaptation for individuals to de-polarize these outcomes.

To my surprise, as I looked into this unlikely connection, previously unexplained facts fell into place in support of this hypothesis, and gradually this thesis took shape. In chapter two, I start with the apparently irrational information processing tendency, known as confirmation bias. I argue that it can play a constructive role in naturalistic group decisions, as might have occurred in a pre-modern evolutionary con-

text. While making collective decisions more reliable, confirmation bias also results in increased polarization of the outcome. This is the starting point for the third chapter, which makes the case that the pattern of probability weighting that is now well documented in the literature, acts to rectify the polarized group assessment. The final chapter then picks up the original question of improving social choice, but from a behavioral perspective. Perfectly rational agents cannot scale their abilities into a procedure for rational group decision making. But behavioral agents, despite their considerable flaws, can structure their interaction in a way that mitigates the logical impasse identified in social choice theory.

Chapter 2

Culture, knowledge and group induction

The day had dawned clear and sunny on Aug. 6, 1945. Sunao Tsuboi, an engineering student at Hiroshima University, was hurrying to class after quickly downing a bowl of porridge and slurping some seaweed soup at a roadside breakfast shack. Okinawa had fallen to American troops, but Mr. Tsuboi doubted that Japan's defeat was imminent. "I firmly believed the emperor was God, and I was ready to die for him," he says.

Suddenly the young man was swept off his feet and hurled 30 feet by a deafening blast - the fury from the first use of an atomic bomb in history. [24]

Overconfidence in collective beliefs can be dangerous.

Beliefs that are shared across a social group may be objectively true or false but have, through some historical process, acquired the status of "social truths" and, taken together, comprise much of the society's culture. How should humans balance their trust in social truths versus potentially contradictory evidence they acquire individually? This depends on how it acquired its status as truth and whether better information than what is available to a single individual was effectively aggregated. If a group process could

be enforced, an information-optimal strategy would have all individuals honestly reveal their private information, which could then be combined through Bayesian statistical inference. Since everyone has this complete information available, they could follow the same Bayesian procedure and reach the same conclusion.

As the reader knows, this procedure rarely occurs in a modern setting and has even less bearing on naturalistic group decision making in a pre-modern setting. The remainder of this paper studies a plausible naturalistic procedure and the implications for the evolution of both group and individual decision making. I argue that in lieu of an explicit, rational aggregation, there is another highly effective strategy for optimal group decision making. But, it has spillover consequences for individual information processing that can explain a number of the most intriguing anomalies in behavioral economics and group dynamics. This chapter will argue that confirmation bias can actually be a useful aid for groups to learn about their world inductively, but has the side effect of contributing to belief polarization. The subsequent chapter shows how this also provides a possible explanation for the phenomenon of probability weighting, as described in prospect theory.

Environment of Evolutionary Adaptation

While human intelligence remains unmatched within the natural world, two qualifiers to consider are, 1) we have only recently recognized the cleverness of some non-human species [29], and 2) our own cleverness relies to a remarkable extent on

knowledge derived from others – we are largely “social thinkers” [73]. From everyday events to fundamental facts that society is based on, the preponderance of our worldly knowledge is derived from the word of others. Our senses tell us that we are hungry, but our food packaging tells us if we will exceed our sodium RDA. A glance out the window may lead us to bundle up for a trip to the store, but a glance at our iPhone now serves just as well.

In the modern context, we have a variety of tools and methods to determine the reliability of information sources, which seems far removed from the paleo-environment in which the parameters and biases of our decision strategies evolved. Rubin [68] refers to the conditions which dominated our recent evolutionary history as the environment of evolutionary adaptation (EEA) (“environment of evolutionary adaptedness” is more commonly used in the evolutionary psychology literature [26, pp.25-56]). If we assume, as Rubin does, that a typical social group during the EEA consisted of 25 - 150 people, then it is plausible that group decisions were made through explicit interactions among the whole population, or a sizable subset with decision making authority. We model group choice as the outcome of a majority rule procedure, through an informal process whereby individuals sequentially make their preferences known.

As an descriptive model for how human groups make decisions, majority rule cannot claim exclusivity. But, the simple majority decision rule has broad experimental support as the default decision procedure in human groups [75, 27, 43, 65, 72]. Social psychology research on group decision suggests that “positions are valued proportional to the level of support they receive” [15]. From an evolutionary perspective, we expect

majority rule to be favored where information is distributed widely, and more hierarchical procedures where relevant information is concentrated at the individual level. Here we focus on the former, in a stylized model to highlight the effect of distributed information. Kameda, et al[44] provide game-theoretic foundations for majority rule under more general circumstances than addressed here.

Robust Group Choice with Heterogeneous Motives

In the scenario we study, we assume that all group members have fundamental preferences for the same (correct) decision. During foraging, for example, all members would prefer the group to choose the direction in which they have the best chance to find food, although members with different information may have different beliefs about which direction is best. However, at the individual level there may be various idiosyncratic or transitory factors influencing preferences, aside from the shared need for sustenance. One individual may have a sore foot, creating an idiosyncratic motive distinct from dietary needs, and prefer the closer destination despite its meager food prospects. Group dynamics may themselves create a more systematic disruption of individual motives, due to family bonds, dominance hierarchies and factions vying for power by building coalitions. We'll consider this scenario where individual motives are driven by the resource utility of finding the best food source, as well as a confounding social utility driven by peer pressure.

Social utility here is the value derived from relationships with others, which

may change depending on the choice taken. Formally, “peer pressure” could be seen as the magnitude of this difference. Psychologically, it is the urge to maximize this value, and includes both affiliative and coercive pressure. Affiliative pressure might come about within a family group or other coalition. For example, if one’s close relations chose an option at odds with one’s private signal, they might be pulled toward that option by internal goals of harmonious relationships, rather than for fear of punishment. This form of peer pressure can exist without another individual incurring disutility.

On the other hand, coercive pressure may involve a direct or implicit threat of retaliation that is potentially costly to enforce. Across primate species there are a wide range of social power structures, from dominant alpha-male (gorillas) to decentralized, female-led panarchy (bonobos). Groups led by singularly dominant individuals may fail to effectively aggregate distributed environmental information. Nevertheless, counterexamples do exist. The olive baboon is generally governed by steep male power hierarchies. But, when deciding direction, the leaders step back and allow a highly democratic process to choose the outcome. For this species, the competition in leadership does not dominate resource selection. In other baboon species, dominant individuals are able to monopolize food resources. These leaders are less concerned with optimizing resource selection, and extend their monopoly on decision to food destination selection [46].

The key point is that there may be many scenarios that produce peer pressure to choose based on criteria other than resource information. Capitulating to this pressure entrusts the decision to a subgroup that, by definition, does not consider all the information available to the group. If all publicly shared messages were known

to be honest, then all members could integrate the information and all make the same, optimal, choice. But, in our model, a much simpler method is also optimal. If each individual votes for (chooses) the direction suggested by their private signal, then majority rule produces the optimal decision rule.

Confirmation Bias and Group Dynamics

Here we introduce confirmation bias into our model, which is a well documented information processing bias at the individual level. Members with an initial signal pointing in one direction, will (on average) exhibit confirmation bias when evaluating evidence from the shared signal information of others. Confirmation bias is the tendency to give greater credence to evidence supporting ones existing belief, and to scrutinize contrasting evidence more thoroughly. It is found to be ubiquitous in human reasoning and is an individual decision bias that contributes to group polarization [60, 71]. It's impact is found in economic decision making [40] and, although the propensity for it varies among individuals, it does *not* correlate with measures of intelligence [85]. The consequences of various forms of confirmation bias have been analyzed [63, 3], but its origins and evolutionary persistence remain unclear. However, a recent proposal by Mercier and colleagues[54] for its evolutionary advantages finds them in group dynamics. In Mercier's model, a behavioral "division of cognitive labor" produces the best arguments for group deliberation, after which group members revise their beliefs by epistemic necessity, after comprehending both information and argumentation.

It is worth pointing out that being "biased" in favor of one's own private signal

is often quite rational. For example, signal veracity is assumed for one's own signal, but might be questioned for the reported signals of others. In this case, one would rationally weight one's own signal more heavily than socially acquired information, creating a "rational bias" in each agent's evaluation in their own signal's direction. The important point is that this occurs for each agent, and neither those who are biased toward the correct answer, nor those who are biased toward the incorrect answer are aware of which it is. Agents are also drawn towards a particular position through coalitional pressure, which has no epistemic foundation. The leader of the coalition has a signal with the same fidelity as the least powerful member. If there is sufficient evidence from socially conveyed signals, then even with confirmation bias the member will revise their opinion, and the group will converge on the correct direction. If the signal strength is weak, then agents will be drawn towards the signal of their coalition partner, which gathers no wisdom from the crowd. Confirmation bias confers more weight to the epistemic (signal) portion of their utility, and less to the conformist coalitional portion, tipping the balance towards the correct answer when the signal strength would not otherwise overwhelm conformist pressure.

As a simplified example, suppose in a group of 33 members there is a dominant coalition of 22 members and a less powerful rival coalition of 11. Signal fidelity is 60%, so the maximum likelihood voting pattern would be 60% correct, and 40% voting erroneously. If the coalition leaders are at peace, there may be no incentive for others to share dishonestly, and the optimal choice will be evident to both leaders. But, if the coalitions are at odds, each coalition leader may attempt to exert their control by

taking the group in the direction given by their private signal. Although the leader of the dominant coalition has 22 members to consult with, revising his position would be tantamount to admitting defeat to the rival leader. So, the leaders are locked in rivalry defending their position, with only a 50/50 chance that the the winner (probably, the leader of the larger coalition) will advocate the best choice.

Components of the Model

N Decision makers choose by majority rule. Individuals vote in a random order, observing the votes of those before them and updating their beliefs, but ignoring the process once their vote is cast.

$\omega \in \{-1, 1\}$ - The state of nature: direction of better resources.

Member i receives signals: s_i^1 (private) from nature, and $\{s_i^k\}, k = 2, \dots, N$ from others.

$\phi \geq .5$ - probability that an individual's private signal is accurate.

$\alpha \in [0, 1]$ - level of trust that a public signal accurately reflects the private information.

Binary choice: which direction should the group hunt/forage.

$\vec{y} = \{y_i\}$ - individual choice/vote. $y_i \in \{-1, 0, 1\}$. $\bar{y}$ = actual group choice.

Utility: $U_i = U_{Si}(\vec{y}) + U_{Ri}(\bar{y}, \omega)$ This model suppresses temporal complexities by having

group members make their (public) decisions in an arbitrary, discrete sequence¹. When it is an individual's turn to choose, they face a static problem, with constants $U_{Si}, U_{Ri} \geq 0$:

$$U_i = U_{Si} * \mathbb{1}_{\{y_i \text{ is socially optimal}\}} + U_{Ri} * \mathbb{1}_{\{\bar{y} \text{ is resource optimal}\}}$$

U_{Si} is the relative (social) utility gain from conforming to net peer pressure.

U_{Ri} is the relative (resource) utility gain from the group choosing the optimal destination.

Honest and accurate social signals: $\alpha = 1$

It could be that there are no factions, coalitional pressures, or other sources of strategic motivation and social information is trustworthy. Individuals update their earlier beliefs with each new public signal. Individual accuracy improves, compared to voting from private signal only, because each voter has access to the “wisdom of the crowd”. In the limit, with $\alpha = 1$, everyone votes identically.

Individual i adds the public signals of others to his private signal of s_i^1 : if there are m observations of s_i^1 and n observations of $-s_i^1$, the odds that this will be observed when $\omega = s_i^1$ versus $\omega = -s_i^1$ is $\phi^m * (1 - \phi)^n$ to $\phi^n * (1 - \phi)^m$, or ϕ^{m-n} to $(1 - \phi)^{m-n}$.

With questionable social signals: $\alpha \in [0, 1)$:

For a social signal with accuracy diminished by $\alpha \in [0, 1)$:

¹Communicative turn-taking is found across all major branches of the primate order, and may be ancestral to their divergence [48].

$$\phi' = (\phi - .5) * \alpha + .5 = \alpha * \phi + .5 * (1 - \alpha)$$

Ex: i observes private s_i^1 , and public signal s_i^2 , and suppose $s_i^2 = s_i^1$. Then
 $P_i(\omega = s_i^1) = \phi * \phi' / (\phi * \phi' + (1 - \phi) * (1 - \phi'))$

With $m+1$ observations of s_i^1 and n observations of $-s_i^1$, the odds that this will be observed when $\omega = s_i^1$ versus $\omega = -s_i^1$ is $\phi * \phi'^m * (1 - \phi')^n$ to $\phi'^n * (1 - \phi) * (1 - \phi')^m$, or $\phi * \phi'^{m-n}$ to $(1 - \phi)(1 - \phi')^{m-n}$.

With confirmation bias

Note that in some cases confirmation bias can be quite extreme, such that even disconfirming evidence is re-interpreted to strengthen the original belief. In this section we consider only the more moderate case. Disconfirming evidence is partially discounted, while confirming evidence is taken at face value.

For a disconfirming social signal with accuracy diminished by $\alpha \in [0, 1]$, $\phi' = (\phi - .5) * \alpha + .5 = \alpha * \phi + .5 * (1 - \alpha)$, while confirming signals maintain accuracy ϕ .

Ex: If $s_i^2 = s_i^1$, then $P_i(\omega = s_i^1) = \phi * \phi / (\phi * \phi + (1 - \phi) * (1 - \phi))$

Ex: If $s_i^2 \neq s_i^1$, then $P_i(\omega = s_i^1) = \phi * (1 - \phi') / (\phi * (1 - \phi') + (1 - \phi) * (\phi'))$

With m observations of s_i^1 and n observations of $-s_i^1$, the (biased) odds that this will be observed when $\omega = s_i^1$ versus $\omega = -s_i^1$ is $\phi^m * (1 - \phi')^n$ to $\phi'^n * (1 - \phi)(1 - \phi')^{m-1}$

Numeric examples with confirmation bias

To get a sense for the effect of confirmation bias, a few numeric examples are worked through below.

Suppose, for example $\phi = .6$, and the member got two confirming and two disconfirming signals after their original private signal.

With honest social signaling and no confirmation bias, the odds for $\omega = s_i^1$ are 6 : 4, unchanged from the original private signal.

With doubtful social signals, $\alpha = .5, \phi' = .55$, the odds are again 6 : 4.

But, with confirmation bias, odds are, $.6^3 * .45^2$ to $.55^2(.4)(.45^2)$, which is about 7 : 4.

And, with maximal confirmation bias (of the moderate kind), $\alpha = 0$, and the (biased) odds are: $.6^3 * .5^2$ to $.5^2(.4)(.5^2)$, which is about 8.6 : 4.

Confirmation bias, with ten social signals, split equally for s_i^1 and $-s_i^1$, the

odds for $\alpha = .5$ and $\alpha = 0$ are approximately: 9 : 4 and 15 : 4, respectively.

Decision to “Defect” or Capitulate to Peer Pressure

U_{Si} is the utility gain from conforming to net peer pressure.

U_{Ri} is the potential gain from choosing the optimal destination, based on private information.

By the standards of individual rationality, a group member should only incur the cost of breaking from their peers’ choice if they believe doing so will improve the chances of the group making the better resource choice. ΔP_i is i ’s belief of the impact their choice will have on the choice of the group decision, $\bar{y}$:

We assume $\Delta P_i > 0$, where $\Delta P_i \equiv P_i(\bar{y} = s_i^1 | \text{defect}) - P_i(\bar{y} = s_i^1 | \text{capitulate})$

Compare utilities:

$$EU_i(\text{defect}) - EU_i(\text{capitulate}) = \Delta P_i (P_i(\omega = s_i^1) - P_i(\omega \neq s_i^1)) U_{Ri} - U_{Si} \stackrel{?}{>} 0$$

$$\text{Defect if } P_i(\omega = s_i^1) > 1/2 \left(\frac{U_{Si}}{U_{Ri} \Delta P_i} + 1 \right)$$

In the special case that $\Delta P_i = 1$, defect if $P_i(\omega = s_i^1) > \frac{U_{Si} + U_{Ri}}{2U_{Ri}}$

When will $\Delta P_i \approx 1$:

- When i is a pivotal voter.

- When i is overconfident of being pivotal or influential, consistent with prior behavioral assumptions. To model this we broaden the role of α , as a behavioral parameter which increases the propensity to vote strictly on the basis of one's private information. In other words, a behavioral self-assessment, $\Delta P'_i$, may be greater than a rational estimate:

$$1 \geq P_i(\bar{y} = s_i^1 | \text{defect}, \alpha) - P_i(\bar{y} = s_i^1 | \text{capitulate}, \alpha) \equiv \Delta P'_i \geq \Delta P_i$$

When is Confirmation Bias Collectively Rational?

Subjective probability of being correct, with confirmation bias:

$$P_i(\omega = s_i^1 | m \text{ signals agree, } n \text{ disagree}) = \frac{\phi^m (1 - \phi')^n}{\phi^m (1 - \phi')^n + \phi'^n (1 - \phi)^m}$$

For $\alpha, \phi' \in (0, 1)$, $\frac{\partial \phi'}{\partial \alpha} > 0$, so $\frac{\partial P_i(*)}{\partial \alpha} < 0$.

In other words, confidence in one's original signal rises monotonically with confirmation bias.

The majority position is the maximum likelihood solution, so it is **always** better (w/r to finding resources) to go with the (private signal) majority rather than a sub-group or particular individual.

“Behavioral choice” is collectively *resource-superior* to rational choice when:

$$P_i(\omega = s_i^1 | \text{bias } \alpha, m \text{ agree, } n \text{ disagree}) > 1/2 \left(\frac{U_{Si}}{U_{Ri} \Delta P'_i} + 1 \right) > P_i(\omega = s_i^1 | s_i^1)$$

And, under some circumstances, when:

$$P_i(\omega = s_i^1 | \text{bias } \alpha, \text{ m agree, n disagree}) > \frac{U_{Si} + U_{Ri}}{2U_{Ri}} > P_i(\omega = s_i^1 | s_i^1)$$

For a single-shot game, confirmation bias is collectively rational but individually irrational. Thus, so far confirmation bias exhibits an evolutionary catch-22, similar to other altruistic traits – it is good for society as a whole but evolution, operating at the individual level, will select against it.

Collective Fitness and Evolutionary Stability

Applying evolutionary reasoning to group behavior can be tricky, since the genes encoding behavioral variants are exposed to direct selection pressure only at the individual level. So, if individual A carries a genetic variant that benefits the group, but at an evolutionary cost to himself (lower probability of surviving offspring), it is the others in the group who receive benefits and have relatively more offspring. Thus in the next generation, there will be a lower proportion of the population carrying the variant. The carrier will be at a relative disadvantage, and if the net individual benefit is negative, the group will eventually have no carriers. While it is possible that the advantage to the group is so great, and the disadvantage to the individual so small, that the gene will increase in frequency. This is the notion of group selection, which is largely discredited because the conditions for this to occur are exceptional – although it is being reconsidered in the context of cultural group identity [66].

Traditional game theory considers the utility of an action to be a subjective assessment made by the individual which, whether motivated by monetary gains or personal goals, is simply an economic ground truth to be taken as a given. On the other hand, evolutionary game theory [84] measures utility in terms of genetic “fitness”, e.g. the extent that the trait increases or decreases the probability of the trait being passed along to future generations through reproduction. While it is convenient to assume, and probably a good first approximation, that individual preferences reflect genetic fitness (a monkey prefers a ripe banana to a fibrous tuber because it is higher in nutrients), it is not generally true. The encoding of preferences reflects past rather than present evolutionary selection.

Evolutionary encodings of preferences can also be idiosyncratic because there is generally not a 1-to-1 mapping of phenotypic traits subject to selection pressure with corresponding genes that can be selected for. Fruit flies that encounter a banana variant that smells and tastes only slightly differently from the others but contains a dangerous toxin, is subject to selection pressure to discriminate this banana variety from others. Since there is no gene that codes specifically for tendency to eat this banana variety, other genes that encode secondary factors will come into play. For example, if the toxic variety grows at higher elevation or emits a distinctive odor, the fly may adapt by favoring lower elevations or avoiding fruits with similar odor signatures, and these adaptations will also affect the fly’s selection of other types of fruits that are not toxic but share elevation or odor traits with the forbidden fruit.

In our model, confirmation bias acts as a evolutionary parameter to modulate

social utility by dampening the impact of peer pressure. To the extent that it is non-specific to the decision making context modeled, it may have idiosyncratic influence on other decision making scenarios, just as a fly’s elevation preference spills over to other food choices. The viability of confirmation bias as a social utility parameter will depend on the totality of its impact on individual fitness. Thus, it becomes more viable if it can be invoked in a more context sensitive way, operating in tandem with other cues that restrict its effect to the scenarios in which it is adaptive. Though beyond the scope of the current model, experimental evidence does indicate that human decision making is highly context sensitive, and the confirmation bias parameter is influenced by multiple factors such as the expectation of a subsequent critical discussion [56] or the relationship to the person who is the source of the information [55]. Mercier and Sperber [54] find that confirmation bias is strong for individuals’ *production* of arguments, but systematically weaker during the *evaluation* of the arguments put forth.

The specific model that we simulate here has no penalty for the individual’s confirmation bias. We only consider the confirmation bias parameter in the context of group decision making. Individuals might behave differently in the context of individual decision making. In fact, that is what I will argue in the next chapter. We will see that in the context of a laboratory experimental economics setting, experimental probabilities that are given to the subject will not be downgraded as less reliable overall, as in group decision making, but will on average be transformed with the probability weighting function experimentally demonstrated by Kahneman and Tversky in 1979 [41], and discussed in 1954 by Wade Hands [9].

Information Aggregation and Decision Variance

Under controlled experimental process, we can test hypotheses about the effectiveness of group decision strategies with great precision, but we are left with the question of external validity. For example, Lorenz, Rauhut, & Kittel (LRK)[51] studied the effect of deliberation in groups of three to nine members, with performance based pay that guaranteed incentive and epistemic compatibility. They came to the surprising conclusion that majority rule underperformed groups that used either a more or less constrained process: consensus decision or allowing group members to choose

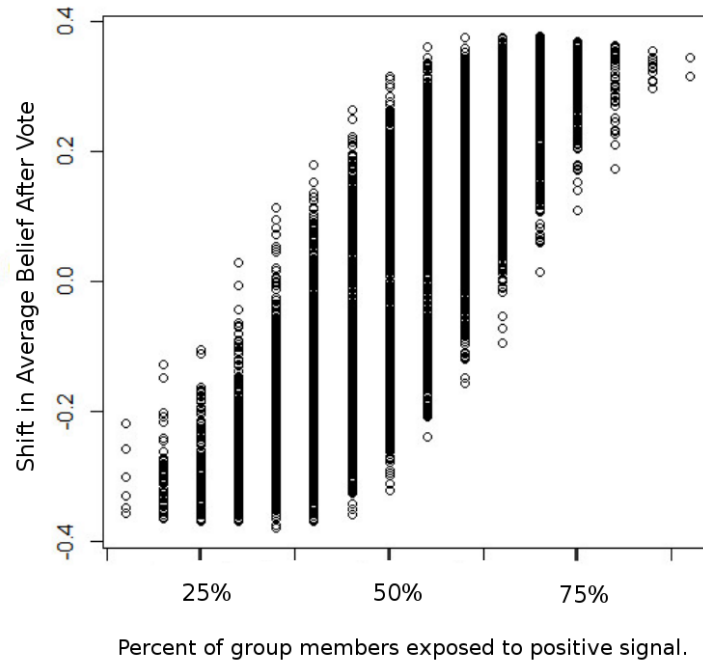


Figure 2.1: Simulations: Zero confirmation bias, high belief variance.

individually. They conclude, “When expert committees are established, they should either decide unanimously or aggregate independent judgments instead of imposing a majority decision.” But, various subtle issues make this conclusion less cut and dry. For example, their group communication protocol used an electronic format that was quite constrained (either a chat window, or strictly limited to numerical values as proposed solutions) and allowed no direct verbal interactions. Recent research has shown that humans react differently to textual versus in person interactions [70]. Even within the majority rule framework, rather than a simple voting scheme, individuals are required to revise their choice until a majority have the same answer, leaving the authors to speculate whether the motives were truly to find the correct answer or to reach an acceptable settlement with other members.

In contrast, a naturalistic decision process must be robust under wide ranging conditions and information aggregation scenarios. Chat rooms are clearly an inappropriate assumption for decisions in the EEA, but more importantly, so is the assumption of uniform dissemination of communicated information. A better assumption is that information sharing would be haphazard prior to decision making. Impressions are built up over time based on information expressed, sometimes in dyad interactions and sometimes in larger group settings where everyone hears what is communicated.

In the simulations presented below, we chose a particular procedure which allows some sources of outcome volatility, but suppresses others. We imagine all the voting members to get together and sequentially state (truthfully) their private signal and then cast their vote. You could imagine casting their vote as stating it out loud,

marking a ballot, or confiding in an elder who tallies the results. In any case, once the person votes, they leave the gathering. This means that the first to cast their vote is not exposed to new information, and leaves with the same belief that they came with. And, the last to vote has the benefit of hearing all the information before casting their vote, and leaves with this revised knowledge. While the last person has the potential for the most dramatic belief revision, if all the shared evidence went against their original signal. It is those in the early and middle stage that have the greatest chance to hear an unlucky sampling of signals that lead their beliefs and thus votes astray. Since there are, in principle, 2^N different information sharing mappings, there are other mappings that would give somewhat different results. But, this iterative voting procedure is simple to describe and illustrative of the impact of confirmation bias.

We will see from these simulations that chance ordering effects matter – who votes first can change the outcome. But, there is another substantial source of variance that we are suppressing with the procedure described: information cascades. Information cascades are a familiar problem in the social learning literature [8], and occur when opinions developed from (partially) shared information are treated as new, independent information. In our simulations, we assume that the signals are shared honestly, so other group members can base their decision purely on the revealed signals. A cascade occurs when group members only know the vote, and not the private signal. In that case, a vote represents a mingling of private knowledge (their signal) and public information (previous votes), which subsequent voters can no longer disentangle. Once the sequence of votes is sufficiently strong to overwhelm the next voter’s private information, all sub-

sequent votes will be for the apparent winning position. Although we don't include this scenario in our simulations, the conclusions below directly extend to cases which include this additional source of error. Including cascades only makes the argument for the value of confirmation bias in this decision context stronger.

Figure 2.1 demonstrates the significant variance in belief change with no confirmation bias. This is not a direct plot of vote outcomes (presented later), but shows the simulation results of net change in group (average) belief after witnessing the voting process. Each point is one simulation run. The X axis shows the net balance of private

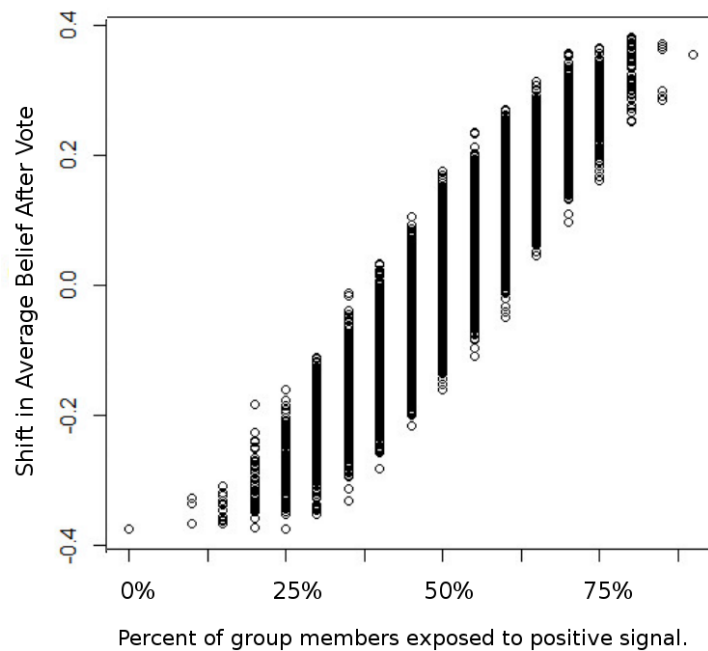


Figure 2.2: Simulations: Higher confirmation bias reduces belief variance.

signals: 0 indicates an equal number of signals, +2 indicates that instead of balanced, one person who received a negative signal is now receiving a positive signal, etc. In these simulations with a population of 20 individuals, the most extreme outcomes are the two runs on the top right, in which all but two individuals receive a positive private signal, so there is a net +16 in private signals.

On the Y axis is the average change in belief, measured in probability. In this set of simulations, the probability that the private signal is correct is .6. even if everyone confirms that belief by placing a vote consistent with their private signal, the most the

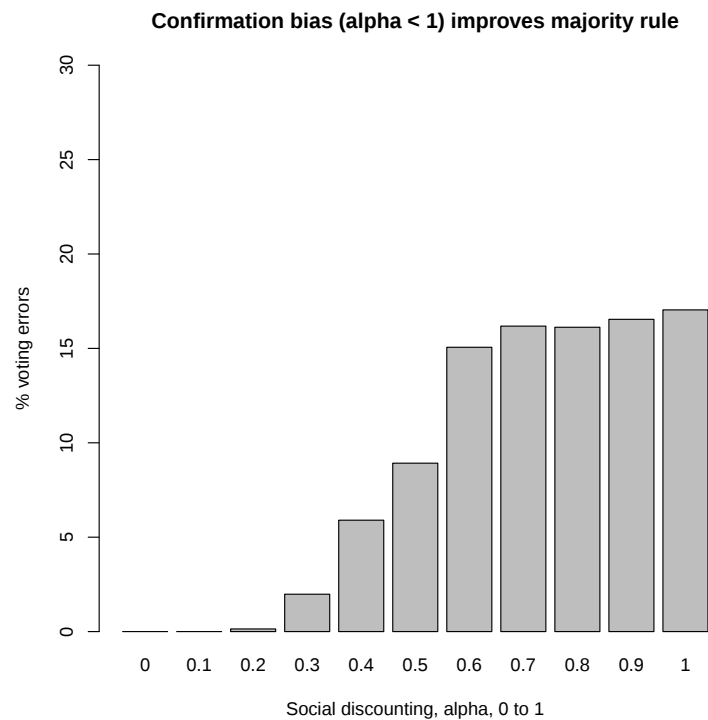


Figure 2.3: Greater confirmation bias (.4) reduces erroneous voting.

probability can increase is .4. Consider the $X = 0$ case, in which there is no objective reason to prefer one direction over another and no rational basis for a net shift in belief. The huge variation is entirely a function of voting order. If a cascade forms early, beliefs will be heavily swayed in that direction. And, if a cascade forms very late or not at all, there will be little net change in belief.

Contrast this with Figure 2.2, in which confirmation bias has been turned on, which discounts opposing information by .4. There is much less belief variance, and correspondingly, majority voting will deliver the correct answer more reliably. The relationship between voting errors due to information cascades and confirmation bias is shown in Figure 2.3.

Behavioral Implications: The Majoritarian Shift

What we have shown so far is an evolutionary model in which individual fitness is consistent with efficient group decision, even though the mechanism involves an apparent distortion of individual rationality. But, figures 2.1 and 2.2 demonstrate another consequence, at the group level. These figures show a change in average belief of the group, away from neutrality and in the direction of the majority opinion. This shift is erratic without confirmation bias, but becomes more consistent and more apparent as confirmation bias increases. Figure 2.4 shows what would happen without the random error introduced from ordering effects in haphazard naturalistic decision making. To generate this figure, we allow each member to hear the private signal information of all

other members, subject to confirmation bias. The X-axis shows the relative balance of signals. When signals are equally balanced, there is no net change – each member will shift towards their own signal due to confirmation bias, but they will be balanced by an equal number moving in the opposite direction. As the proportion of private signals increases in the, say, positive direction, more members beliefs will drift in the positive direction and with greater confidence, since they will be exposed to more confirming signals. This exhibits the classic pattern of group polarization, in which a group with an imbalanced belief distribution will become more extreme through discussion and decision making.

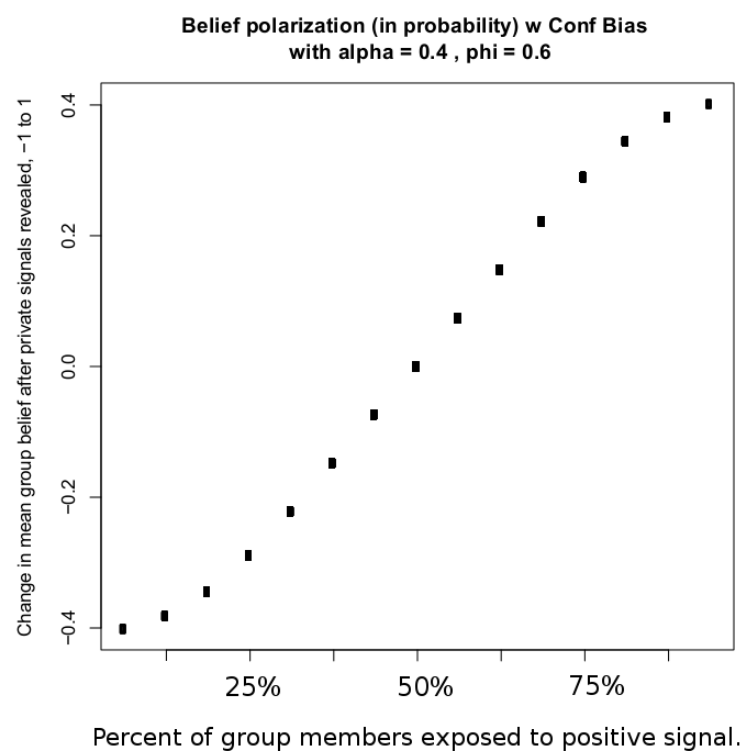


Figure 2.4: Simulations w/o noise: Confirmation bias induces group belief polarization.

Group polarization driven by the majoritarian shift

The phenomenon now widely referred to as “group polarization” was first noted by J.A. Stoner in the 1960’s, when he experimentally tested the prevailing wisdom that groups would exert a moderating influence on individual judgment. To the surprise of many, he found the opposite, and termed the effect the “risky shift” in group decision making [77]. Experimental participants showed an increased willingness to support risky behavior in the context of group decision making, relative to their individual preferences prior to participation in the group decision. This turned out to be an instance of a more general tendency of groups to produce more extreme results in multiple domains, whether it is in taking risks or in judgments of probability. In the domain of jury awards, it has been termed the “severity shift” [69], and was named a “satisfaction escalation” in marketing literature [12]. This phenomena is so ubiquitous [59, 5] and general to group dynamics it has been dubbed, in an eponymous journal article, “The law of group polarization” [78].

It is important to note that “polarization” here does not refer to bifurcating one group into two polar camps. On the contrary, this refers to a consistent empirical shift in the average of individual positions. When there is a range of options for a given decision or group probability evaluation, groups of individuals that on average already tend away from the ambiguous middle, after group discussion, will more decisively (as an average) move away from the middle and toward a pole. Thus, in contrast to Stoner’s hypothesis, this dynamic will tend to bias groups, and individuals after

group deliberation, toward a more extreme action. There have been many attempts to explain group polarization through psychological mechanisms, each showing merit within particular domains, but this is the first proposal of an adaptive evolutionary mechanism that might contribute to this phenomena across the board. In this light, the risky shift or group polarization, etc, are individually descriptive, but “majoritarian shift” provides a causal terminology that is agnostic to whether risk, caution, or some other attribute is at issue, and indicates the likely direction and magnitude of the shift rather than just that it is towards polarization.

Chapter 3

The social induction theory of probability weighting

“Correlated Variation” – I mean by this expression that the whole organization is so tied together during its growth and development, that when slight variations in any one part occur, and are accumulated through natural selection, other parts become modified. This is a very important subject, most imperfectly understood, and no doubt wholly different classes of facts may be here easily confounded together.

- Charles Darwin, The Origin of Species [17, p155].

There and Back Again: Connecting the behavioral biases.

The central argument of this chapter is that several of the most enduring puzzles of behavioral decision making could well be closely related, and that viewing them as working synchronistically offers a plausible causal explanation that remains elusive when considered individually. I also argue that the apparently inscrutable nature of these biases is a result of the profound difference between modern settings in which they appear irrational, and the environment in which they mutually evolved.

The previous chapter developed a model in which confirmation bias led to better information aggregation for group decision making. One by-product of this was that individuals, on average, developed more extreme opinions than what is merited by the data available to them. The results of this model are consistent with a well documented phenomenon, often referred to as “group polarization”. We proposed to use the more informative descriptor, “the majoritarian shift”, but may use the terms here interchangeably to emphasize either the majority influence or the polarizing result.

We also introduced the concept of the EEA (Environment of Evolutionary Adaptation) in the previous chapter, positing that behaviors or biases that appear irrational today should be evaluated by their adaptedness to the environment in which they evolved. This chapter will develop the concept of an EEA majoritarian process and how that differs from contemporary democratic processes. We argue that what we think of as “cultural knowledge” represents implicit group decisions about what is true, and this decision process carries the inevitable mark of belief polarization as developed earlier: the naturalistic, majoritarian decision process of human groups is a powerful tool for inductive knowledge aggregation, but tends to distort probabilistic beliefs towards the extreme.

Finally, we argue that this systematic belief polarization of cultural knowledge is compensated at the individual level. What has been seen as one of the most dramatic and consistent violations of rational probabilistic decision making, known as “probability weighting” in prospect theory, becomes a rational individual adjustment when viewed through the lens of the EEA. This presents a coherent theoretical model,

grounded in evolutionary theory, that ties together and explains the persistent enigmas of confirmation bias, group polarization, and probability weighting.

Behavioral decision making and probabilistic choice

The paradigmatic economic agent makes decisions in accordance with expected utility theory: in the absence of certainty, the utility of a possible outcome is proportional to its value as well as the probability of its occurrence. Sometimes the probabilities are considered to be known, such as when rolling dice or playing a well defined lottery in an experimental economics lab, and otherwise the agent must make a subjective probability assessment consistent with Bayesian reasoning. Daniel Kahneman and Amos Tversky pioneered the experimental evaluation of this paradigm, and while they found substantial overlap with the rational agent model, they also found persistent and systematic violations [41].

The violations of expected utility they observed involved apparent inconsistencies with both the calculation of value for the prospective outcome, as well as the incorporation of probability to construct its expected value. The descriptive theory Kahneman and Tversky developed, to provide a more accurate model of human choice behavior, was called “Prospect Theory” [41], and their later refinement to correct particular failings that were discovered over time was called “Cumulative Prospect Theory” [42].

Figure 3.1 shows the characteristic shape of the probability weighting function that Kahneman and Tversky found in a series of experiments for their 1992 paper. This

is a refinement from the shape they originally proposed in their 1979 paper, and is broadly consistent with the work of many other experimentalists (with some notable exceptions explored below) before and since this paper was published. Their original idea was that people, for some unknown reason, did not use probabilities directly in their calculation of value. In experimental settings where probabilities for particular events were provided unambiguously, subjects behaved *as if* the probabilities had first been modified by a weighting function, and then used for the calculation of expected value.

The graph in figure 3.1 shows that probabilities near or just below the center, as well as the very extremes of 0 and 1, are treated as we expect. But, probabilities near the extreme values of 0 or 1, are mapped to significantly less extreme values. For example, a probability of .1 might be mapped to $W(.1) = .15$, and .9 might be mapped to .7. In their original formulation of prospect theory, probabilities were weighted independently from the value of the prospect. As can be seen in the figure, the later version allowed for distinct weighting functions when dealing with losses, as opposed to gains, and had a somewhat more complicated implementation via the cumulative distribution function rather than the probability distribution function. Since our theory does not diverge from expected utility with respect to the value function, we will keep the original prospect theory as our frame of reference and look in more depth at the probability weighting function in that context, and we will utilize the most widely used functional form as developed by Drazen Prelec [62]. The weighting transformation for probability p is given

by:

$$\omega(p) = \exp(-\beta(-\ln(p))^\alpha), \text{ with } 0 < \alpha < 1 \text{ and } \beta > 0$$

The two-parameter model allows control over the two noticeable features that identify the vast majority of probability weighting shapes that have turned up in the experimental literature, both in aggregate and individually. The one we focus on is the degree of curvature of the inverted S-shape, controlled by parameter α . Parameter, β , shifts the function up and down, which might correspond to optimism and pessimism or other anchoring effects.

EEA model of majoritarian decision

How should one model group decisions in the EEA? The model should reflect broad patterns consistent with experimental evidence, without assuming specific procedures or modern conveniences – so, no ballot boxes, per se. A survey of the research literature shows the enormous scope and subtlety of group decision making with, for example, the term “group decision making” returning in excess of 250,000 articles in the academic literature, as cataloged by Google Scholar. Two widely published veterans in this field, Norbert Kerr and Scott Tindale, enumerated several paradigmatic functional forms to represent group choice from the social psychology literature as shown in figure 3.2 [45].

The functions represent a probabilistic mapping from individual preferences to group choice, given some sort of group decision process. The X-axis represents the probability that a given member of the group would make the correct choice, in a binary

choice setting. The Y-axis gives the probability that the group will choose correctly. The diagonal line is simply a linear, or null model, or can be thought of as the prediction for a group of size one. The other lines show how the mapping is influenced by the problem domain. The uppermost curve represents a choice problem in which discerning the correct answer might be hard, but once presented with the correct answer, it can be easily verified. If a small minority of the group has discerned the right answer, they can demonstrate it to the rest of the group, and the whole group will choose

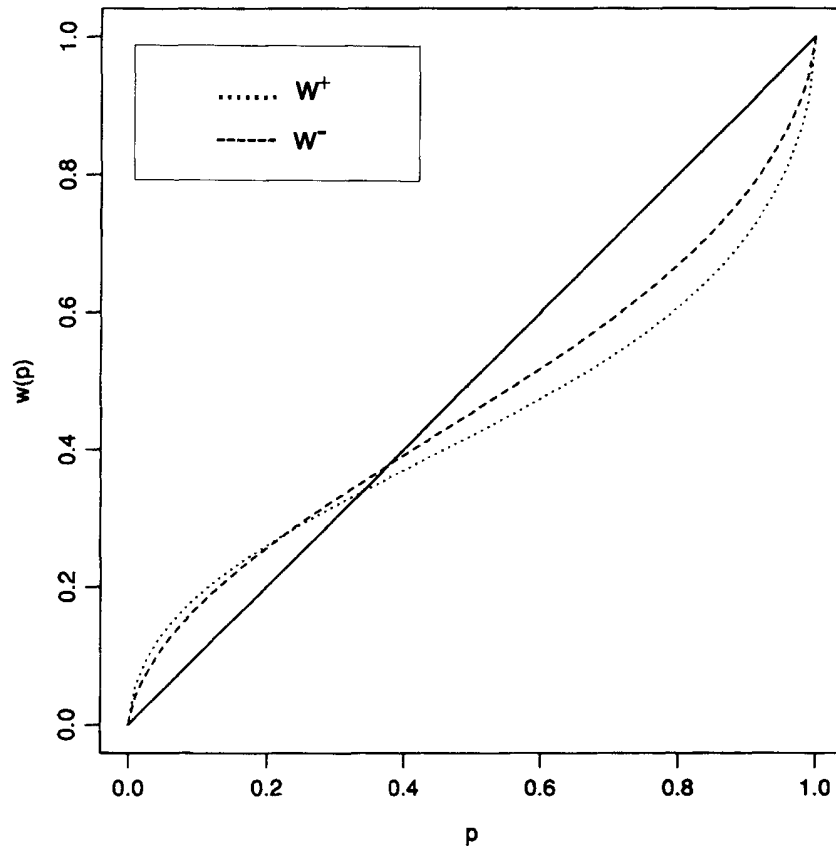


Figure 3.1: Kahneman & Tversky (1992) Probability Weighting Function, W^+ for gains, and W^- for losses.

based on a full revision of their beliefs. They call this a “truth wins” decision scheme. Other decision problems fall into an intermediate category, in which demonstration of the solution helps others understand that it is correct, but only to a degree, denoted the “truth supported” decision scheme. These are classes of problems that have been studied extensively, and which Kerr and Tindale characterize in more detail and provide references to the literature, but our interest is with the final curve.

The third curve depicts the majority rule function and corresponds to those decisions which remain dependent on cultural knowledge. Keep in mind that in this particular context, beliefs and decisions are not entirely distinct. A cultural belief is one that is shared and instilled within a given group or tribe. And, we are describing

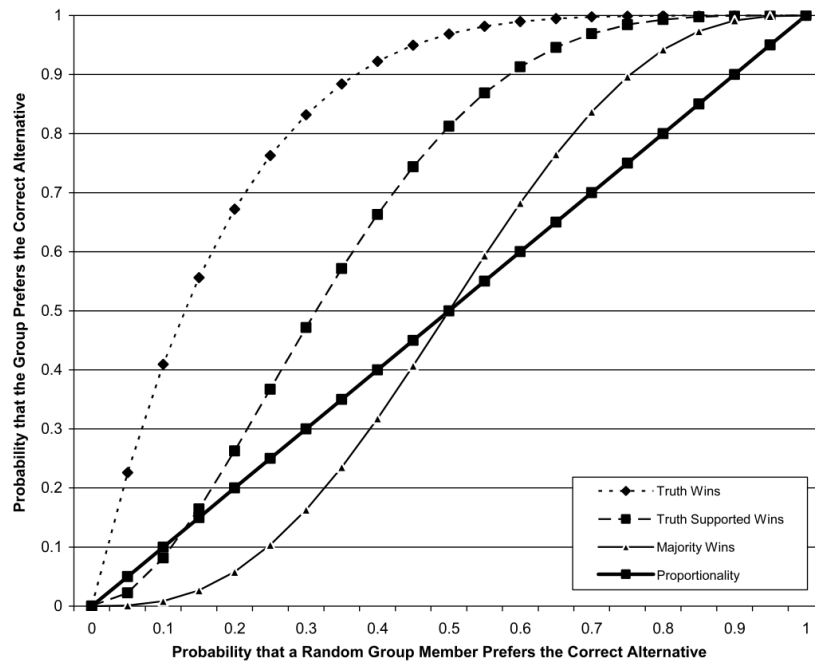


Figure 3.2: Kerr & Tindale (2011) models of group performance

a decision process, probably implicit, by which the belief is adopted. Beliefs which are eventually confirmed or disproven might still be considered part of “the culture”, but their resolution by assessing the state of nature removes or reduces their causal dependence on the group process.

The majority function might apply to the mythical aspects of culture – elaborate beliefs in the afterlife, for example, which may or may not provide indirect material benefit – but we would have to define what the “correct alternative” meant in this case. But, more aligned with the rest of this thesis, the majority function would most clearly apply to those cultural beliefs which provide useful knowledge of material reality and are learnable inductively, but are not practicably amenable to deductive solution. Whether rabbits or an important herb could probably be found on the far side of a given mountain is an inductive problem. Cultural beliefs can hold the current state of knowledge, but a specific probability could never be confirmed simply by inspection – it is learned over time through experience, and subject to change in the long run.

For the remainder of the chapter, we will be concerned only with inductive group beliefs, and consider how an individual should interpret those decisions, after the fact, once the group decision process is considered settled.

Effective Condorcet size

Figure 3.3 illustrates different levels of confidence that might be gained from aggregating knowledge by a majoritarian process. This is based on the well known “Condorcet jury theorem”, in which there are N people with independent signals about

some value of interest. As the number of people voting on a proposition grows, from three to thirteen in figure 3.3, the probability that the majority will judge the value correctly increases. It is remarkable that this is true even when each person's signal is mostly noise, and they have only slightly better than chance in judging the proposition individually. In fact, as long as the probability of the signal being correct is slightly greater than 50-50, with a large enough group the majority winner will almost always be correct.

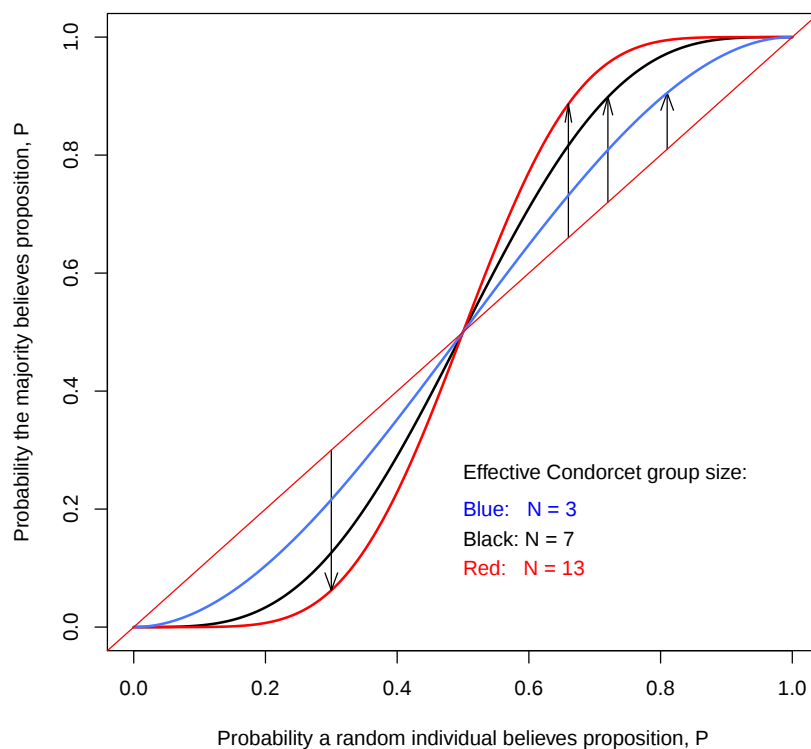


Figure 3.3: Majoritarian shift from individual to group belief.

The catch is that information is rarely, and probably only briefly, independent. In more typical situations where information is highly dependent (circulating on social media, or propagating along familial ties in the tribe), adding more people does not provide as much an improvement in discrimination. We capture that issue here by defining an “effective Condorcet group size”. Bray et al [13] used a similar concept, “functional group size”, to rate a groups’ productivity, based on the idea that as group size increases, more members simply stop participating in group problem solving. Effective Condorcet group size is a measure of how much independent information the group has available to it. A group of twenty people with highly interdependent information sources may only have the information content of three people with independent signals. A majoritarian process in this group could do no better than a group with effective Condorcet size of three. This is important because individuals must make a subjective judgment of how confident they are in the cultural belief. , as discussed below, humans seem quite bad at estimating correlation between the information received by different individuals. So, a group of twenty who underestimate the true correlation might be driven by a majoritarian process with effective size of seven, thus imparting a stronger majoritarian shift than is warranted by the true weight of aggregate information. Also, note that since “effective group size” is really a measure of information content and not a body count, it is also plausible to consider non-integral values.

On the other hand, someone attempting to utilize a belief that has previously been determined through a majoritarian process faces the reverse problem: How many individuals actually participated in the decision, and how independent were their infor-

mation sources? As an example, if four group members gather independent information about the environment and mention it to several friends. The next year, the same issue comes up, and all of those friends talk about what they understand to their other friends, and then the whole group makes a decision. At best, their effective Condorcet size will be four, despite dozens participating in the decision.

For our prehistoric forebears, learning about rare environmental events (the black swans of ecological niche variance) is an extremely valuable ability for temporally continuous cooperative living groups. The lifespan of any individual is not sufficient to sample these rare events. For a group that relies on fishing in the adjacent river, it is important to know the range of extreme flooding events when choosing a settlement. A farming community needs to not just maximize annual production, but also must minimize the chance of catastrophic crop failure. The critical factor in evolution is not short term growth, so much as long term survival. Before there were clinical trials, this information was passed down through cultural knowledge. It is well accepted that oral tradition can preserve knowledge of substantive historical events for 500 to 800 years, and now there is growing evidence for cultural memory extending more than 7,000 years[61]. But, knowledge from one event, passed through many voices in the next generation, does not gain statistical significance just because many people attest to it. It would have been highly advantageous for the prodigal hominid to build some intuitive mechanisms for affording weight to social information in a way that was sensitive to its origin. Prior probabilities are how we think of this today. Different categories of information sources do not necessarily require different prior probabilities, but lacking

evidence to the contrary, it is a reasonable assumption.

“Culture” and probability weighting: from EEA to experimental economics

The term “culture” has many layers of meaning, depending on the context in which it is used and the interests of those using it. For our purposes, culture should be thought of in an anthropological context, in the environment of evolutionary adaptation. As such, it might comprise all thoughts, beliefs and values that derive from the common knowledge and practices of an individual’s tribal society, in contrast to those derived from the individual’s personal experience. But, we will only be concerned with the belief component, and will ignore culturally instilled values, which we would not be able to do if we wanted to understand the origin of an individual’s complete utility function.

If an individual samples an unknown fruit and finds it to be distasteful, they have gained personal knowledge and can use that to avoid the fruit in the future. Was the fruit simply unripe or a “bad apple”, and the avoidance might deserve to be tested again in the future? The confidence in the probabilistic belief in the fruit’s terrible nature will be weighed against the individual’s other knowledge to decide on complete avoidance or cautious monitoring of the fruit’s value.

Separating knowledge obtained individually from that derived from cultural beliefs is not generally straightforward, as they may become blended over time with the best decision coming from an appropriate merging of the evidence. We consider cultural knowledge to be the component of an individual’s beliefs that originated from

social interchange with other members of the tribal society, rather than through the individual's direct experience. This may come from direct instruction, as is commonplace in modern schools, or through knowledge diffusion by overhearing and observing other group members.

The critical difference between cultural knowledge and individually acquired knowledge is that culture was the product of a collective decision process – not a formal democratic voting scheme, but an informal and, we will assume given the discussion above, majoritarian process. Cultural knowledge is that which has gone through a decision process such that it is collectively agreed to be true, and subject to whatever distorting influences that process imprints on it. Without knowledge of how a specific item of cultural knowledge came to be, the most likely process was majoritarian, and the most likely imprint left on it is belief polarization.

We suggest that, over thousands to millions of years, humans evolved to process two kinds of knowledge differently: individually acquired knowledge and group level, cultural knowledge. Cultural beliefs, like individual beliefs, may sometimes be considered true or false, but generally will be probabilistic and comes in degrees of belief, or confidence. For example, kinship relations may be known with certainty, but the chance of finding a specific herb on a hilltop is only considered likely. Probabilistic beliefs that are formed from primary individual experience can be taken at face value. Probabilistic beliefs that are derived through a polarizing majoritarian process must be de-polarized before using them in individual choice. We suggest that probability weighting is an automatic, non-conscious process built into the hardware of human probabilistic assess-

ment, which co-evolved with collective decision making in human groups to de-polarize group beliefs for more accurate use by the individual.

What we now understand as probability has subtly played a dual role in human history through its interplay with group process, and evolution has accommodated this role with mental processing that seems peculiar and misguided in the modern context. Prior to the scientific method, there was no formal system for testing theories and establishing truth. Statements of facts, in antiquity, were the product of cultural negotiation and assimilation, through a collective dynamic which tends to generate biased outcomes. On average, collective judgments are shifted to be more extreme and imbued with greater confidence than rational analysis would permit. In an economics experiment, when a subject is told that the probability of winning a given gamble is 10%, it is stated as a known fact and not as the opinion of the experimenter. The brain processes this as a socially derived truth. They are asked to perform a task, with

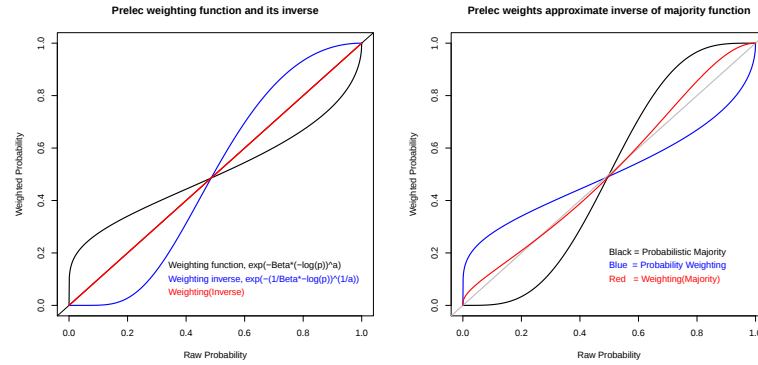


Figure 3.4: Probability weighting inverts majoritarian polarization.

the knowledge that the established, agreed upon, fact calls for a 10% probability value. The appropriate brain hardware to process this statement of “fact” is the one called upon to assess social truths that had been settled through collective dynamics, rather than through scientific experimentation or calibrated instrumentation. This does not negate individual judgment as to which “facts” are plausible and which deserve to be handled with suspicion. Instead, it is positing that our Bayesian prior, in making that assessment, is weighted towards the lessons of evolutionary history.

Figure 3.4 shows the approximate inverse relationship between the majoritarian and a hypothetical probability weighting functions. This comparison is using the Prelec weighting function, which is not to say that the Prelec function as the “actual” Prospect theory weighting function. This appears to be the most often used functional form in the literature, but the detailed shape of Prelec’s function was determined by axiomatic as opposed to evolutionary considerations. Thus, we have no reason to expect them to be exact inverses. The first image graphs the probability weighting function with its exact functional inverse. The red line shows the function $y = \text{weighting}(\text{inverse}(x))$ which, as expected, is the diagonal. Figure 3 shows the majority function, the weighting function and $y = \text{weighting}(\text{majority}(x))$, which is close, but not quite, diagonal.

If our hypothesis is correct, that social probability is treated specially, then we should look for a dividing line between social probability processing and individual probability processing. However, it isn’t entirely clear where the evolutionary line would be drawn. Would the line be drawn between what one’s buddy said the score was

at the big game the night before, versus what a group of friends say it was, or is it a neural processing distinction that goes much deeper than that? By identifying the EEA, we've clearly pushed the boundary back in time, and wouldn't expect any evolutionary adaptation to the specifics of sports betting behavior. But, evidence of the inverted-S shaped probability weighting in Macaque monkeys [76] points to potentially a more ancient origin than we typically think of for the EEA. Distinct neural processing between social and individual probability information may, in fact, predate language and symbolic thought.

Experimental economic and psychology experiments have not been honed to look for this dividing line. But, in the methodical pursuit of understanding the nature of probability weighting, just such a distinction was found, and has become a topic of increasing scrutiny. It is known as the “description-experience gap” in probability weighting [34]. Probabilities, when provided to subjects as statements of socially uncontested truth, were (on average) de-polarized by probability weighting. Probabilities that were uncovered through direct experience were not. The following sections review the evidence.

Consistency with experimental findings

Expected utility theory abstracts away all conscious processing, but of course real brains have processing stages and modules that are mostly unconscious. These ultimately get integrated and synthesized into conscious control in the brain's frontal

lobes, which manage top level executive functions. Our standard economic formalisms only have one kind of probability, and a unitary consciousness with which to process it. On the other hand, the theory developed here suggests that from an evolutionary or neuro-biological perspective, we may think of two varieties of probability: that derived directly from individual experience and that derived from a dynamical social decision process and comprising a shared set of cultural beliefs. The gold standard of scientific evidence is from generating novel and unexpected predictions from a model and then finding them verified by experiment. Evolutionary psychology has been subject to criticism for providing post hoc explanations that are difficult or impossible to test experimentally. One can counter this argument with appeals to Ockham's razor, but without delving into the deep questions and literature involved we can say that at a minimum, as with cosmology, the theory should comport with known facts. In the case of human probability processing, not only do the facts seem surprisingly consistent with the theory, the theory provides a possible explanation for another outstanding puzzle in the experimental literature: the so-called description-experience gap.

Individual heterogeneity with distinct sub-types

A striking aspect of experimentally measured probability weighting functions is the overall consistency of their shape within aggregate data, while demonstrating comparatively wild variation at the individual level [30]. The simplest explanation, that it makes little real-world difference within the range of observed variability, is disputed by a growing number of studies showing real impact in economic decision making [21, 7]

as well as with critical non-monetary life choices such as medical decision making [11]. Furthermore, this variation is not primarily due to individual random error, because behavioral patterns at the individual level are relatively stable [28].

Finally, these stable individual patterns fall into distinct subtypes, the most well-established being between “classically rational” expected utility maximizers and the pattern established for prospect theory and seen in aggregate data, such as figure 3.5 from Fehr-Duda and Epper [21]. They used hierarchical Bayesian models to determine the best categorizations into subtypes, which are shown in the figure. The red curves represent the majority group, comprising approximately 80% of the population in both data sets, and blue represents the remaining 20% of the population. Graph a shows the results for students from two Swiss universities, and replicated earlier results from other Swiss and Chinese university populations. Graph b is from the Swiss general population, and the analysis again found a roughly 20% subpopulation that was significantly different. But, in this case the difference was primarily in the elevation or degree of optimism, and the authors concluded that there is likely a continuity of degrees of probability distortion in the general population.

There are many intrinsic factors that seem to systematically impact the weighting function, such as gender and age [22, 81], as well as distinct subtypes that appear within exceptional populations (universities). While we are beginning to map variations in probability processing to measurable differences in neural imaging (discussed below), we do not yet have direct evidence of inheritable genetic traits that specifically control the probability weighting function. But, other aspects of risk processing are known

to be regulated by specific genes [89, 90, 16]. The correlation of distinctive risk processing to different life stages and societal roles is not proof, but is at least suggestive, that selection pressure has tuned probability weights differentially, in response to the specific needs of these situations. This argues that probability weighting, rather than being a fluke or error or simply sloppy evolutionary adaptation, is a finely tunable tool of evolutionary responsiveness. Diversity of weighting profiles may simply reflect the wide array of decision scenarios that the human genepool has been responding to. The urban intelligentsia sampled at universities may display higher proportions of a heretofore disadvantageous individualistic rationality, which is no longer subject to the strict cultural confines of a single tribal identity.

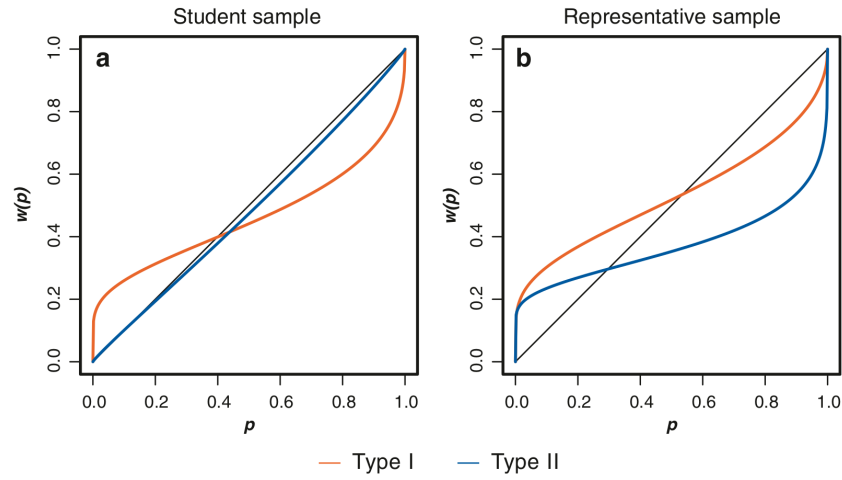


Figure 3.5: Population analysis reveals non-random heterogeneity

The “description-experience gap”

The development of prospect theory by Kahneman and Tversky provided a breakthrough in cross-fertilization between economics and psychology, and contributed to the meteoric rise of experimental economics as a discipline unto its own. Thus, the framework of how to experimentally analyze probability was largely seeded by their early work. This typically had experimentalists convey to the subjects a written statement of what the outcome probabilities would be in a given gamble. This was thought of as an objective probability by the experimentalist, because they could cause their tools to generate whatever probability distribution they required.

But, by the late 1990’s researchers began to stumble upon experimental designs that did not reproduce the well established “inverted-S” probability weighting. In fact, they were getting results that were nearly linear and often actually reversed: an S-shaped weighting function. The key difference in the experimental protocol was that the experimentalists did not describe to the subjects what the probabilities were in advance. They required the subject to try different options and learn directly what the probability distribution looked like from experience [6]. They had discovered the “description-experience gap”.

Potential methodological problems caused some doubt initially, due to the challenges of working with more extreme probability values. To test weighting of very low probabilities by experience, the subject must be exposed to a very large number of samples simply to construct an accurate probability estimate [33]. Substantial method-

ological refinement seems to have ruled out misleading subsampling [25] or recency effects [82], although there is still debate as to whether decisions from experience produce S-shaped or linear weighting functions [14, 35], or whether other factors mediate this variation in results.

Finally, experimentalists have devised another, ingenious method for testing the differences between described probability (i.e., social truth) and inherently individ-

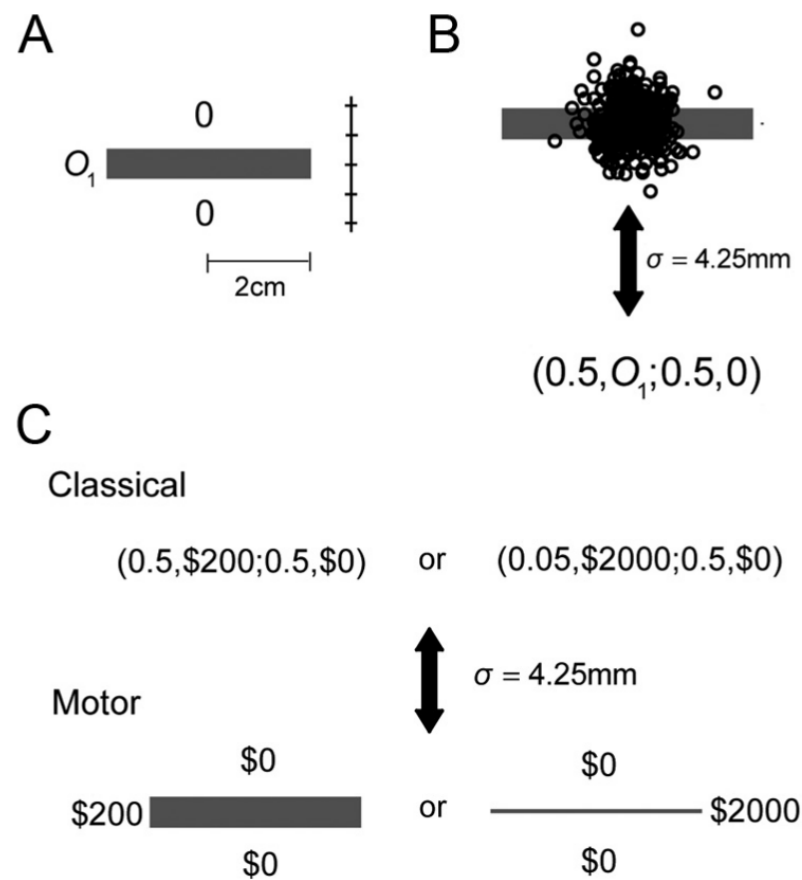


Figure 3.6: A motor lottery uses the subject's own error distribution to construct gambles

ual probability – by using the uncertainty of the body’s own motor skills [53]. Figure 3.6 illustrates the elements used in Wu, Delgado and Maloney [86, 87], and is of a similar structure to other motor lotteries in the literature. Part A of figure 3.6 shows a sample target that would be briefly displayed on a touch screen monitor. The subject in the experiment is trained to attempt to touch the target with their index finger, but the target is small enough and appears briefly enough so that the subject will often miss. After a training period, the size and speed are calibrated so that for a given probability, a target of the correct difficulty can be generated such that the subject will have that likelihood of hitting it. Part large B of the figure shows a target that had been calibrated to be hit 50% of the time, showing the distribution of attempts to strike the target. This is the motor equivalence of the classical lottery just below it, for some payoff, O_1 . Now the system is capable of generating a motor lottery choice that is equivalent to any classical lottery choice. Part C shows an example of such equivalency: the subject has a 50% chance of striking the thick bar and winning \$200, whereas on the right the bar is much thinner, corresponding to a 5% chance of hitting the bar and winning \$2000. Motor lotteries have also shown a linear to S shaped weighting function in aggregate data, and substantial variation at the individual level.

Results of Wu et al [86, 87] can be seen in figure 3.7, showing estimated weighting functions for both classical descriptive probability assessment, and equivalent motor lotteries. The thick, dark curves show the median response, and demonstrate a significant difference between the classical inverted-S and the S-shaped weighting function found in motor lottery studies. Once again we see significant variation at the individual

level, but the median and aggregate results support the social induction hypothesis for descriptive probability weighting.

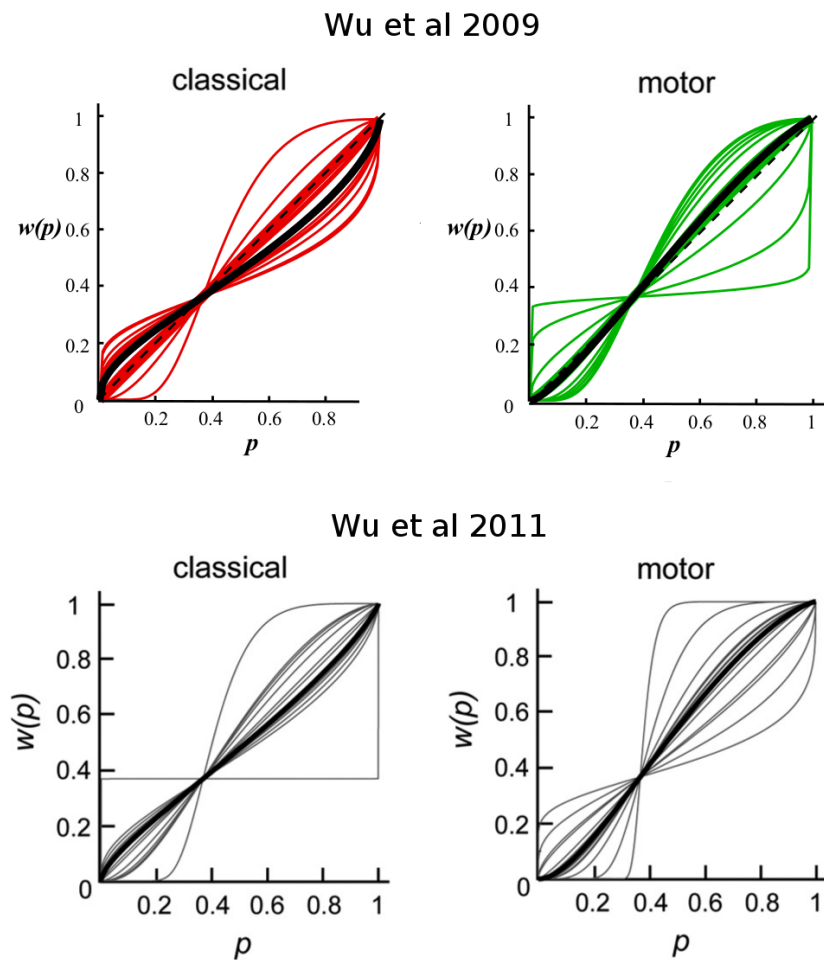


Figure 3.7: Classical descriptive lottery weighting functions on the left, and equivalent motor lotteries of the same subjects on the right. Dark curve is median response.

Neuroimaging evidence

Earlier in this chapter we suggested that to test this theory, we should look for a dividing line between social probability processing and individual probability processing. But, it wasn't clear what level of granularity evolution would choose to define the partitioning. Do we expect a gamble against a group of peers for who will win a sports game should be weighted, while a gamble against an individual would not? Or, is it a more ancient primate instinct, predating language and relying on more basic task information?

An additional tool that may shed some light on this is function magnetic resonance imaging (fMRI). With fMRI, we can see which parts of the brain are engaged when a specific function is being performed. Figure 3.8 depicts the brain from the side, facing to the right, showing color coded regions associated with with probability weighting in the Wu et al [87] study. The data is generated by having the subjects perform the

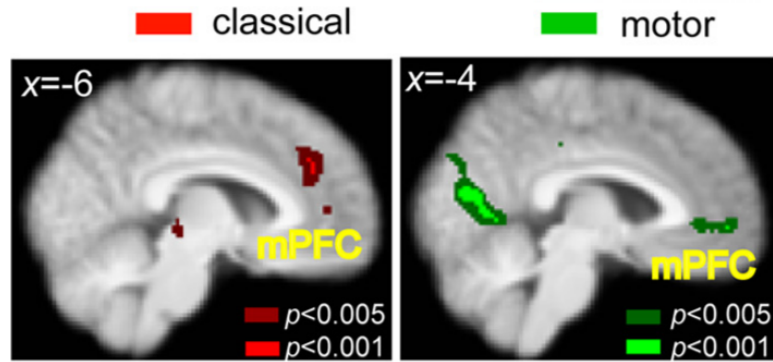


Figure 3.8: Relatively little overlap is seen between regions correlated with probability weighting on the classical versus motor tasks.

classical and motor lottery experiments while their brains are being imaged by an fMRI machine, which measures metabolic activity as a surrogate for neural processing. The colored regions overlayed on the brain images shows those regions that were statistically correlated with degree of probability weighting across the set of individuals in the study. So, this is revealing aggregate patterns and obscures from view the substantial variability in brain regions recruited by different individuals in the study. The two tasks both recruit regions of the medial prefrontal cortex (“mPFC” in the image). The motor task recruits areas that are more ventral and includes part of the posterior cingulate cortex, as well as the large green zone in the back of the image, part of the visual cortex. On the other hand, the classical task recruits more dorsal areas of the prefrontal cortex, which is consistent with past neuroimaging studies of descriptive probability weighting, particularly the dorsolateral prefrontal cortex (dlPFC) discussed below. As can be seen, there is relatively little overlap between the motor and classical tasks. Let’s consider what this might mean.

Tobler et al [80] performed a neuroimaging study to address the question of whether probability weighting occurs at the time of decision, or if it was encoded at an earlier stage. They “induced value” onto a set of meaningless symbols through standard reinforcement training. Symbols were shown on a screen and each one was associated with an unchanging probability and value, which were described explicitly to subjects ahead of the exercise. During the experiment, subjects were shown the symbols while being scanned with fMRI. Each time a subject was exposed to the symbol, they would win that value (or not) at the stated probability, yet no decision was required. Tobler

et al found that post-training the subjects had acquired a preference for the symbols roughly proportional to their expected values – except the expressed preferences were more accurately modeled by prospect value, using probability weights. Furthermore, the subjects who showed the typical inverted-S descriptive weights showed proportionally more activation in the dorsolateral prefrontal cortex (dlPFC). And, given the high variability at the individual level, his subject pool included some individuals with the less common S-shaped probability weighting. These individuals showed proportionally more activation in the ventrolateral prefrontal cortex (vlPFC).

Note that the ventral portion of the PFC was also identified with an S-shaped distortion in the motor task of Wu et al, discussed above. Rather than probability distortion being an artifact or degradation of signal through imperfect neural processes, it appears that the brain is well equipped to tune the level and direction of probability distortion as it sees fit, by neural interconnections with these brain centers. However, that doesn't seem to translate into identical usage of this toolkit across individuals. The significant diversity at the individual level could reflect a number of possible causes. Perhaps the selection pressure is weak or in flux. Or, it could reflect the modern mixing of populations that had each been more finely tuned for interactions with closely related clan-mates. From an adaptationist perspective, it makes sense to have the functional resources that can be flexibly applied to different evolutionary conditions through selective pressure, if these opposing functions are both needed but in differing degrees depending on context. Just as many modern cultures are comprised of people with many skin colors, no longer serving a finely tuned service to the local solar index

and vitamin D needs, we might find that the diverse utilization of these brain regions becomes more tightly clustered when looking at genetically isolated, intact ancestral populations.

Another striking fact stands out from the Tobler et al experiment. Figure 3.9 shows a change in activation of the dlPFC and vlPFC regions over the duration of the experiment. The light bar on the left is a measure of the non-linear distortion from activation of the dlPFC during the first half of the training sequence of the experiment. This corresponds to what we know from standard prospect theory and descriptive probability weighting: participants were told what the probability of these reward events were, and the brain weights, or “corrects”, the values. The theory as argued here is that, on average, inductive cultural knowledge is exaggerated toward the extremes, and the best results come from slightly correcting that for use individually. As the experiment

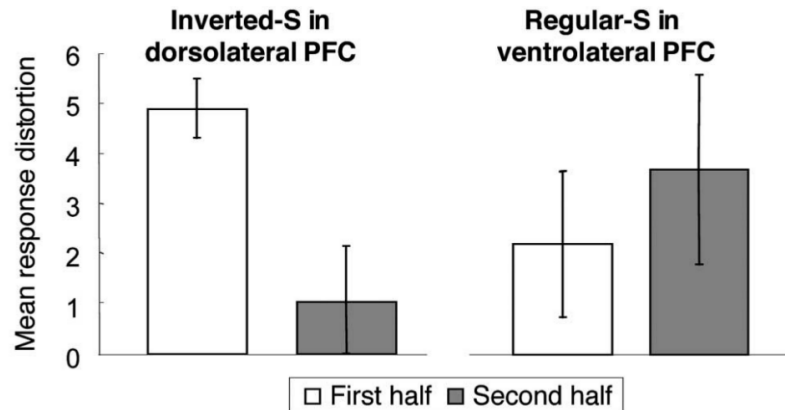


Figure 3.9: dlPFC responds strongly in the first half of the experiment, in response to descriptive probabilities, then less in the second half after learning from experience has occurred.

progresses, the subject experiences the probabilities directly, confirming the original information. The dark bar on the left shows that during the second half of the training sequence, dlPFC has significantly dampened the non-linear rectification of the probability estimate. On the right, we see the suggestion of a reverse process for the subset of individuals that started out with an S-shaped distortion from vlPFC activation, but it is not statistically significant.

These results are consistent with a world where most decisions are based on the merger of somewhat polarized cultural beliefs, augmented with individual experience. Initial cultural knowledge is conditioned by the dlPFC, which is gradually reversed through direct experience which activates the vlPFC. Laboratory experiments using “pure” cultural knowledge (decisions from description) would demonstrate an inverted-S bias. Laboratory experiments in which subjects reasoned with “pure” experiential data, would on average see an S-shaped bias. And in fact, this is what the experimental literature shows.

Discussion

To put this into context it is helpful to have an idea of the overall role, or what we know of it, that the dlPFC plays in cognition. As part of the prefrontal cortex, it is involved in the highest level of executive function, such as error detection, behavioral control, and task-switching [37], as well as what is known as “theory of mind”. In order for humans to function well in an intensively social environment, they must track not

only their own preferences and goals, but must also have a sense of the preferences and goals of others around them. The dlPFC has been found to be critical to understanding the intention of others [83], for assessing and telling lies [38], and its activity scaled with the degree of being misled by the other agent [18]. dlPFC encodes variable needed for strategic choice in economic games [4], when to punish an unfair offer in the Ultimatum game [47], and other moral assessments involving self or others [31]. Players of the 'stag hunt' game who were better able to coordinate and had ability to model higher levels of interactive reasoning ("level k" thinking), had increased activation of the dlPFC [88]. Closer to home, the dlPFC helps us to adapt our own risk preferences by learning from the behavior of others [79]. And, of course, the dlPFC helps us integrate knowledge of the prior probability distribution into a revised understanding [64].

But the dlPFC is far from the only region that helps integrate prior probability information. In one study looking for such regions [32], participants were given a visual shape categorization task, along with descriptive prior probability information about which shapes they were likely to observe. The regions identified in this study as facilitating the use of prior probabilities were (see figure 3.10) the dorsolateral prefrontal cortex (AKA, middle frontal gyrus), inferior frontal junction, anterior insula, inferior parietal lobule, intraparietal sulcus, head of the caudate, posterior cingulate cortex, and fusiform gyrus. But, the dlPFC and vlPFC seem to perform a unique, coordinated tuning function that conditions the prior probability distribution before use, based on context. The key point is that of all the brain centers identified with integrating prior information, the dlPFC is exceptional in both inducing the distortions

noted in prospect theory, and also being a brain center noted for evaluating social and interpersonal intentions and truthfulness. The theory presented here argues that this is not by coincidence.

Hsu et al have hypothesized that an adaptationist account of how evolution shaped the brain is more likely to come from understanding how cognitive components map to prospect theory, rather than expected utility (EU). This is because “prospect theory follows from psychophysics and EU from normative logic. Establishing a neural and evolutionary basis of prospect theory could provide an illustrative example of how the foundation for principles guiding social science might be usefully shifted from relying largely on logic, to respecting biological implementation” [36]. Time will tell if this theory developed in this chapter stands up under further scrutiny. But, as the explana-

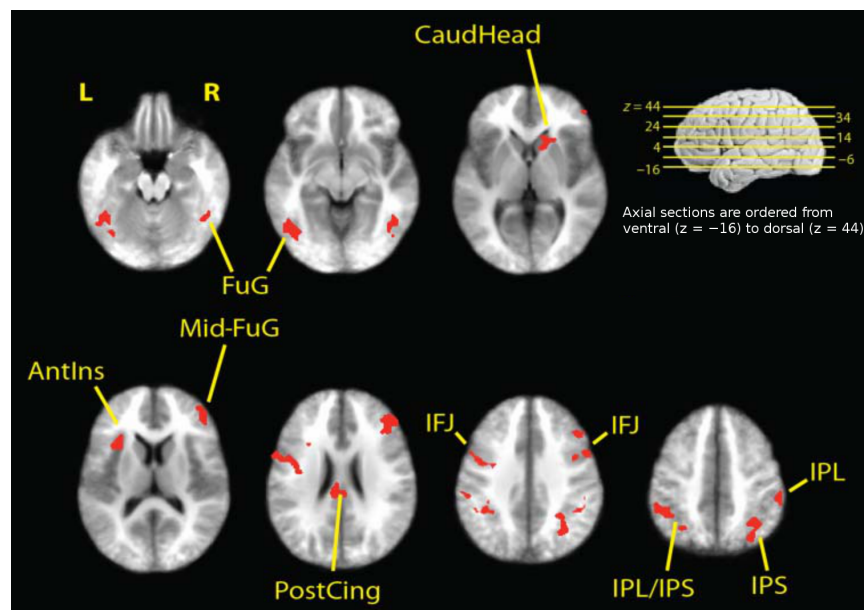


Figure 3.10: Quite a few regions in the brain have been linked to integration of prior probability information. From Hansen, Hillenbrand and Ungerleider (2012)

tory power of unadorned logic reaches its limits, the foundations of economic behavior will necessarily draw more closely to cognitive science, bound together by evolutionary theory.

Chapter 4

Social Choice in the Wild

Deliberative decision making and group decision stability

Kenneth Arrow's impossibility theorem in social choice theory is remarkable for being at once both definitive and inscrutable. It seems to disallow a rational procedure for collective choice without suggesting practical ways to reduce the probability or magnitude of the problem. To shed light on this issue, and considering the difficulties of recreating real-world political dynamics in the lab, we study a pressing social dilemma in a coastal California city. Local citizens engaged in small group deliberations over strategies to resolve an intensifying water supply shortage – an issue which had provoked sharply divided public opinion. We propose a measure of decision stability to grade the danger of a social choice paradox and find that the process of small group deliberation significantly increases this measure of stability.

Social choice theory and the democratic process

Individual rationality plays a central role in economic theory. The paradigm of utility maximization leads to conclusions about behavior only to the extent that agents are actually capable of understanding and consistently taking action to achieve preferred outcomes. One of the standard assumptions of what constitutes rationality is transitivity

of preference: if an individual prefers option A over option B and prefers B over option C, then they necessarily prefer A over C. Social choice theory examines whether this notion of rationality can be scaled up to decision making by groups. Majority voting is one method groups might use to make a social choice, but the theory considers all possible decision making procedures, aside from the dictatorial case of always following a particular individual's preference.

Given that a group's members are individually rational, is there a way to aggregate their preferences while guaranteeing that group decisions meet the same standard of rationality assumed for its individual members? Starting from seemingly very reasonable assumptions, the startling answer provided by Kenneth Arrow was that it is not [1]. Any method used for aggregating a social choice is prone to violating the very basic characteristic of rational judgment described above, transitivity of preference. An important strand of conservative political theory drew on this result to argue for minimal government [67]. Given that the theorems of welfare economics insure that market-based solutions at least achieve Pareto optimal allocations, it was argued that even the most well intentioned democratic system would be capricious and manipulable, in contrast. Some saw this as a potentially fundamental challenge to the legitimacy of democratic governance [50].

On the opposite end of the spectrum are those that see this as motivation for deeper citizen engagement through deliberative democracy. Rather than a passive process of aggregating intact preferences, group dialog and democratic engagement is seen as the opportunity to transform preferences based on shared insight [58]. As

deliberative and participatory approaches to governance have entered the mainstream [39], interest in the more arcane relation to social choice theory has also grown [19]. Ultimately, this is an empirical question.

Experiments in "Deliberative Polling"

James Fishkin has championed the concept of a deliberative opinion poll as a nimble corollary to deliberative democracy [23]. Rather than requiring deep public engagement on every issue, governing bodies can gain insight into public preferences on complicated and contentious issues by having a randomized sample of constituents participate in a deliberative dialog. Their post-deliberation preferences attempt to provide a representative sample of what public opinion would be if there was sufficient opportunity and time for more extensive participation. Farrar, et al [20] have provided some of the first empirical evidence on social choice by deliberative process. To understand their result, first let us look at a potential escape-hatch from the impossibility theorem. One of the assumptions of Arrow's theorem is that any decision making procedure should operate successfully no matter what the "input", as given by the profile of individual preferences that comprise the group. This is sometimes referred to as the assumption of "impartial culture", and implies that no a-priori structure can be assumed to exist in the probability distributions of the population's preferences. In fact, Duncan Black first pointed out one kind of structure that eliminates the paradox in his study of committee voting, called "single peakedness" [10].

Imagine laying out, along a line, the set of alternatives that are being considered. For a given individual, imagine placing their most favored alternative down first. Then, on one side you place the second most valued alternatives and on the other side place the third alternative, alternating sides until every alternative has been placed. For this individual there is a single preference peak, and in either direction, preferences for the options decline. Now imagine adding another individual. If the line is not single-peaked for the second individual (the peak may be at a different location than the other individual, but still declining in both directions), it is possible that a reordering of the options can be found such that the line is single peaked for both individuals. This line is referred to as the structuring dimension. If there is a structuring dimension that is single peaked for all individuals, then there will be a (non-dictatorial) decision making procedure that preserves group preference transitivity.

Farrar, et al[20] and List, et al [49] argue that deliberation increases collective structure in the distribution of preferences, such that groups move closer to single-peakedness. Their measure is global, in the sense that it looks at changes across the entire population of deliberators. On the other hand, we don't learn what happens between individuals within the same deliberating group. An additional limitation to their measure is that while single-peakedness is a sufficient condition for transitive group preferences, it is not a necessary condition. The method proposed in this paper addresses some of these concerns.

Water, public good and debating desalination

One of the challenges of modern democracy is the sheer complexity of some of our societal choices. As we have overcome some of the past threats to our well being through science and technology, inevitably, the remaining or developing threats contain additional elements that render a techno-fix more problematic. Deliberative engagement may be useful, but potentially may require new approaches for 1) issues that require a conceptually demanding analysis, and 2) issues that are driven largely by emotion. The topic of this study demanded both complex information assimilation as well as processing of emotional content: should residents of Santa Cruz, Ca. build a costly sea water desalination plant to augment the water supply?

Residents were concerned by the cost of the plant as well as its energy consumption and other potential environmental impacts. But they had to weigh this against declining aquifers threatened with contamination from encroaching seawater. There was also increasing vulnerability to drought, tightened federal restrictions on accessing rivers and streams in an effort to resuscitate native fish species, and complex hydrogeological considerations that had already launched more than one PhD dissertation.

Though water is a necessity, only a fraction of what we use is a critical necessity and the rest supports a lifestyle and consumptive choices that Americans are accustomed to making freely. But water from desalination is about ten times more expensive than existing water sources, with the cost of building the plant well over \$100 million. Due to the cost and increasingly entrenched viewpoints from public authori-

ties and groups advocating for or against the plant, desalination also proved to be an emotionally charged issue.

Presenting the different perspectives

After studying the water supply and desalination issue, reading local news coverage and responses left by readers in the newspapers' online forums, and going to a number of debates between representatives of the pro and con perspectives, I partitioned the issue along five dimensions. These represented five independent lines of argument that had been made for or against building the plant: two in favor, two opposed, and one that was positioned as neutral which urged further research into legal and regional factors bearing on the proposal. I developed a ten page Issue Guide for participants to review and discuss in a small group setting. The five perspectives formed the architecture for introducing the most salient facts and pertinent values that had been argued in the actual ongoing debate. I fact-checked the points with published research or the public record (water department research reports, newspaper accounts of problems at established desalination plants, etc) and had the document reviewed by members of the water agency, advocates of desalination, and leaders from anti-desalination organizations.

The key factual points were introduced (roughly) in equal proportion among the five sections, with each section arguing its case from the values laden perspective of an advocate of that view. In this way, readers had exposure to both core facts and the

most salient values that had arisen in the literature and debates on the topic. During small group meetings, we took turns with participants reading a paragraph or so of the issue guide and then taking a moment to discuss or clarify what was covered before rotating to the next person. The design of the reference material in this fashion was intended to mitigate common problems that have been noted in group discussion.

In the natural flow of group discussion participants look at each fact with respect to the ultimate question at hand – whether to build the desalination plant, in this case. Typically, this draws individuals into expressing their opinion and arguing their perspective from the beginning, before even the basic facts have been laid out. Biased information processing, the preferential acquisition of information that supports one’s preconceived beliefs, is particularly strong when one anticipates having to defend one’s self [57]. Another documented issue with deliberation is that small groups are not likely to have a balanced representation of diverse views, or equally important, an even balance of argumentative charisma. The tendency for groups with an initial disposition in favor of one view is to follow what has been dubbed the “law of group polarization” [78] and become even more disposed towards the original bias. For these reasons, the group was instructed to focus the discussion on the five positions presented in the guide. They were not forbidden from expressing their own views (and inevitably they did) but the balance of time was spent looking at the strengths and weaknesses of the constructed positions rather than staking out and defending their own views.

Survey on seawater desalination and representative values

An initial survey was conducted in person, with a team going door-to-door over several neighborhoods in the Santa Cruz and Soquel Creek water districts. Neighborhoods were chosen to sample the socio-economic range of the region. The response rate was close to 50% of households contacted, and this was the first self-selection challenge in getting a truly representative sample. The second and more substantial challenge was finding enough individuals from the initial survey group who were willing to join one or more deliberative sessions. Participants were paid roughly \$30 per hour for two hour small group meetings or \$100 for a three hour large group meeting.

The survey used in the door-to-door sample was given again, without modifications, when the participant arrived at their first group session and at the end of every group session. We elicited their level of support for the desalination plant itself as well as for a hypothetical list of citizen-candidates for a desalination task force. In this scenario, task force members would make a recommendation on the desalination plant after conducting an in-depth, independent review of the facts. For the initial survey, participants were only provided with a one sentence summary from each candidate, expressing that candidate's initial attitude about the plant (see the survey in Table 4.2, below). These attitudes had been selected to correspond to the most prevalent "camps" in the public debates. The survey ratings for different candidates were driven by the respondent's own views towards the candidate summary positions and also modeled the gradual revelation of candidate information that occurs in representative elections.

Survey Response	Group Attendance
483	door-to-door surveys of town residents
37	survey respondents attend a small group
24	others recruited to small groups
61	total number of attendees in small groups
31	small group attendees also attend big group
18	survey respondents attend big group only
49	total number of attendees at big group

Table 4.1: Number of people and level of participation.

There were eight small group discussion sessions in which group members read aloud and discussed an Issue Guide which provided an overview of the water supply problem and then presented the five positions of the hypothetical candidates in more detail. Each position was motivated by a primary value orientation that was independent of the other perspectives. Thus, for example, it would be possible for a respondent to highly support the candidacy of all five prospective committee members , to indicate that they highly valued having each value orientation represented on the panel (this did occur but was atypical.) The body of argument for each candidate was roughly one page and contained core facts that pertained to the final decision, so that by the end of the five arguments participants had engaged with most of the central issues.

There was also one large group that everyone was invited to which featured presentations from four experts – two advocating in favor of the plant and two advocating in opposition. The participants then broke into smaller groups for discussion and to generate questions for the presenters. And finally, the large group was reconvened with an open discussion between presenters and attendees. The breakdown of how many

Desalination Plant Survey

Name or ID _____

Survey # _____

The Santa Cruz and Soquel Creek water districts are considering building a seawater desalination plant to supplement our water supply. If a committee of citizens was established to judge whether or not to build the proposed desalination plant, what viewpoints would you be most comfortable with for the citizens sitting on that committee?

Each candidate seeking the committee position honestly explains their values and existing opinions about the desalination plant. But they also vow to take in new information by intensively studying the topic along with other committee members and revise their views in order to find the best solution for the community at large. If the following is all you hear from the candidate, how strongly would you favor or oppose them joining the committee:

Candidate 1:

It is important to maintain our quality of life and our local economic vitality.

(+) Initial attitude: favors the desalination plant.

Candidate 2:

We need to live within our means through conservation.

(-) Initial attitude: opposes the desalination plant.

Candidate 3:

First address the the bigger picture: agricultural water use and overly complex water laws.

(?) Initial attitude: can't decide on desalination until the bigger picture is better understood.

Candidate 4:

Diversify our water supply for a more sustainable and flexible water management.

(+) Initial attitude: favors the desalination plant.

Candidate 5. Consider global warming – desalination is energy intensive.

(-) Initial attitude: opposes the desalination plant.

100 = Strongly Favor

50 = Don't care, either way.

0 = Strongly Oppose

Table 4.2: Description of candidates given in the survey.

participated in which part of the experiment is given in Table 4.1. (Unfortunately, scheduling issues prevented many of the attendees from joining both sessions and the results reported here are focused mostly on the initial small group discussions.)

Preferences for desalination before group deliberation

The question which is typically of primary interest in a deliberative polling event concerns what difference deliberation made on the final preferences for the actual topic of deliberation. A growing literature argues that deliberation has a substantive and salutary impact on policy preferences [2, 52]. Making this case is not entirely straightforward since there is no objectively correct reference point for preference outcomes. The data from this experiment does provide some support for these arguments from the initial pattern of preference adjustment, rather than its final direction.

On the left side of Figure 4.1 is a kernel density estimation followed by a boxplot of the distribution of levels of support for the desalination plant in the initial sample of all 483 households. While the median voter is almost exactly ambivalent, we see that the range of feelings span the gamut but there are distinct spikes of individuals with strong feelings against and in favor of the proposal.

The plots on the right reveal a quite striking result. For all those people who came to a small-group deliberation, these show their preference distributions from when they were surveyed door-to-door (on top) and the distribution from an identical survey when they arrived at their first deliberation (below). Note that there was no contact

with or information provided to these participants between these two surveys. The top right plot looks quite similar to the overall initial distribution on its left. It is still bimodal but has a somewhat larger representation of the desalination skeptics. But the lower image shows that in the time between their initial off-the-cuff survey at home and when they arrived at their first group discussion, their preference distribution has substantially shifted – towards neutrality.

This was unexpected and I can only speculate on plausible explanations. The first hypothesis considers that a large majority of the initial survey population was relatively uninformed on this issue. Even though it had been generating ongoing press coverage and community organizing, most people were familiar with only a small percentage of the relevant facts. Some of these people registered a neutral vote but most gave a gut reaction that was either yea or nay, depending on how they saw it fitting into their framework of values and political orientation. Once they made a commitment to attend a deliberative event to learn about and discuss the topic, they began to follow the issue more closely through discussions with friends and/or reading news accounts. Considering the level of complexity that is actually involved, by the time they arrived at the deliberation event they had an appreciation of their relative ignorance and could no longer fit the question neatly in an established corner of their political values framework. If this hypothesis is correct, the data reflects favorably on the flexibility of average citizens to reappraise their judgment and acknowledge humility. It also speaks to the inadequacy of standard telephone or in-person surveys which the deliberative polling approach is trying to remedy. On the other hand, when participants came to the de-

liberative event they knew that the goal was to discuss the issue with an open mind. Thus, an alternate explanation is that they simply reported their opinion to be more neutral than it actually was, because they felt that they were expected to be initially neutral. The observed effect is likely to be a combination of these factors.

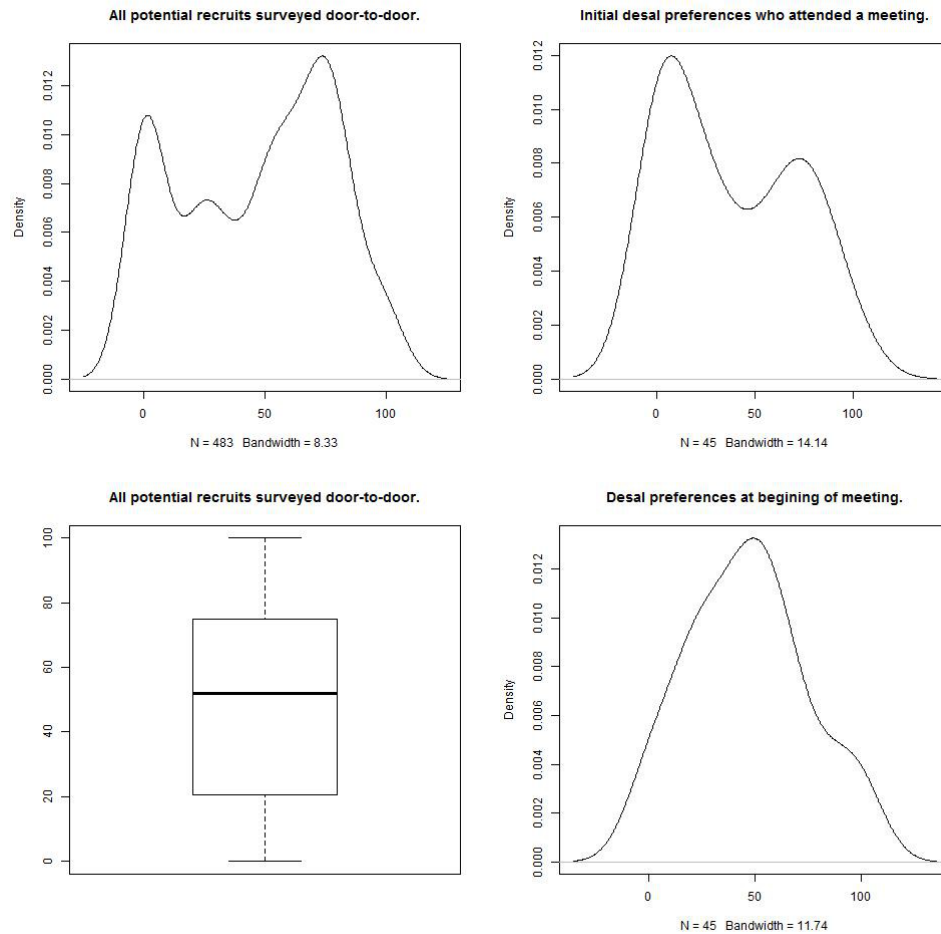


Figure 4.1: Desal Preferences: Left: All door-to-door. Right: The ones who attended

Impact of deliberation on preference for building a desalination plant

Figure 4.2 shows the before and after density maps for individuals attending the small group (left side) and those individuals who only attended the big group (right side). Note that the before, or "pre", preferences shown here are from surveys at the

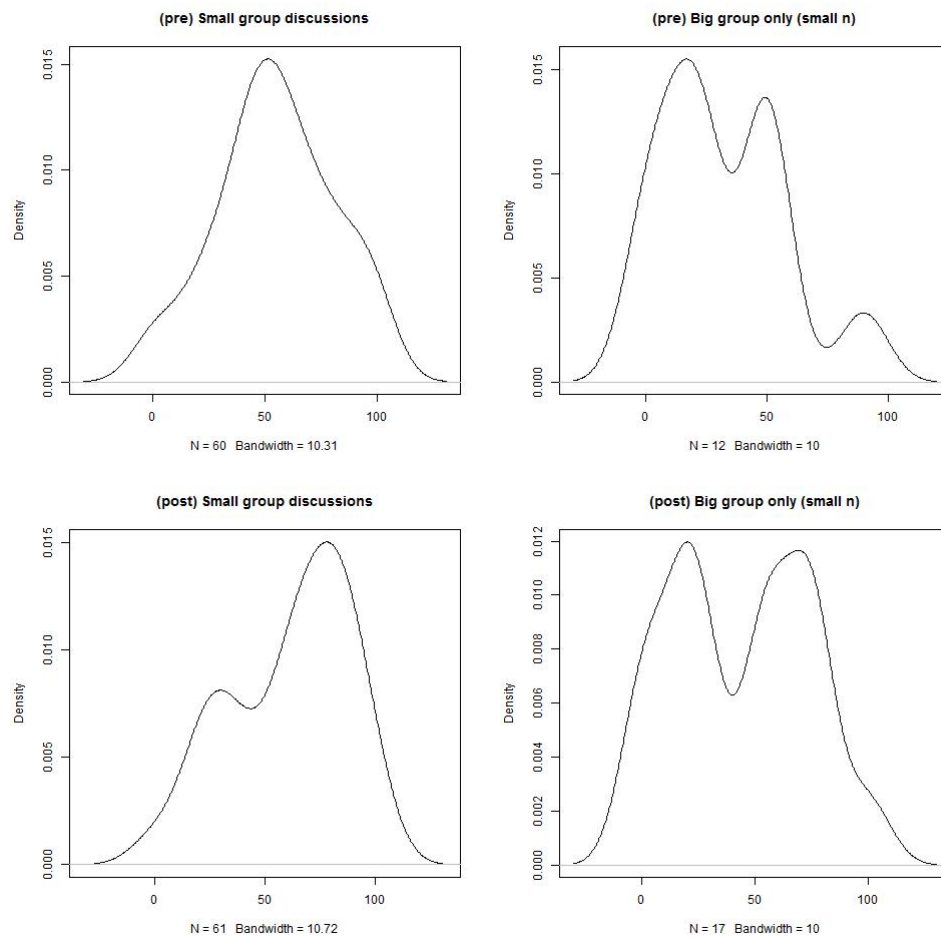


Figure 4.2: Desal preference shift: small group (left) and big group only (right).

beginning of the meetings, and the earlier door-to-door surveys are no longer considered. Also, recognize that the right hand side should be considered cautiously because of the relatively small number who only made it to the large group session and because a number of these failed to turn in the pre-deliberation survey. But the distributions on the right are consistent with the trend shown on the left, that as people learn more about the issue, they are more willing to make the investment and pay for the electricity required to build and operate the desalination plant.

Finally, Figure 4.3 shows that the trend continues: those who attended both a small group meeting and a large group meeting have shifted the most in favor of the desal plant. The left side shows the density plot for this group and the right side shows the trend in box plots from 1)pre-small group, to 2)post-small group and then 3)post small and large groups.

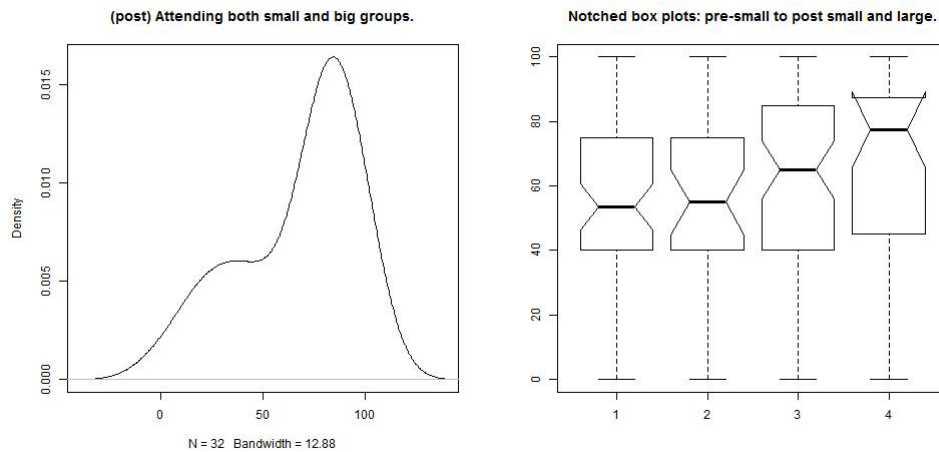


Figure 4.3: Left: Desal preference density of those who went to both small and large. Right: 1)Pre-small group, 2) Pre-small who made it through big group as well. 3) Post-small group, 4) Post small and big for those who did both.

Structure out of chaos: an exploratory analysis

Although the shift in preferences is important, it is not the primary question addressed in this paper. Deliberative democracy theorists have argued that the dialog process itself may generate a shared structural correlation between the different dimensions of the preference space. The increasing correlation eliminates the viability of the "impartial culture" assumption of Arrow's theorem.

The first approach I took to assess the changing correlation structure is through linear regression. If correlations between different policy questions increases, the information-theoretic content of the preference profiles should become smaller and some policy issues should become more predictive of others. Given the breakdown of the survey into the five candidates plus the issue of the desalination plant, it was natural to look at whether preferences for the candidates become more predictive of preferences for the plant.

The first regression, in Table 4.3 below, looks at preferences from the initial door-to-door survey. Due to the relatively large number of respondents, we are able to identify relatively weak correlations, with an R-squared of around .3. This indicates that there are pre-existing structural correlations among these components in the population at large. This is already at odds with a pure notion of "impartial culture".

Table 4.4 shows the results from surveys just prior to small group discussions. As noted above, participants adjusted their opinions about desalination between their first encounters with the door-to-door survey and actually coming to a group discussion. But there is very little change with respect to overall predictive correlation with

	Estimate	Std. Error	t value	$Pr(> t)$	
(Intercept)	28.19875	3.90358	7.224	2.01e-12	***
c1	0.25595	0.04123	6.208	1.17e-09	***
c2	-0.11872	0.04329	-2.743	0.00633	**
c3	-0.03947	0.03709	-1.064	0.28777	
c4	0.32432	0.03778	8.584	1.2e-16	***
c5	-0.10302	0.04063	-2.535	0.01155	*
FORMULA: $\text{desal} \sim c1 + c2 + c3 + c4 + c5$					
—					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 26.68 on 477 degrees of freedom					
Multiple R-squared: 0.3059, Adjusted R-squared: 0.2987					
F-statistic: 42.05 on 5 and 477 DF, p-value: 1.2e-16					

Table 4.3: (Door to door survey) Preferences regressed: desalination versus candidates.

	Estimate	Std. Error	t value	$Pr(> t)$	
(Intercept)	40.82716	16.56830	2.464	0.01695	*
c1	0.38119	0.14011	2.721	0.00875	**
c2	-0.20704	0.15075	-1.373	0.17532	
c3	0.01498	0.13024	0.115	0.90884	
c4	0.12073	0.14323	0.843	0.40301	
c5	-0.07036	0.15111	-0.466	0.64335	
FORMULA: $\text{desal} \sim c1 + c2 + c3 + c4 + c5$					
—					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 22.23 on 54 degrees of freedom					
Multiple R-squared: 0.3306, Adjusted R-squared: 0.2686					
F-statistic: 5.333 on 5 and 54 DF, p-value: 0.0004696					

Table 4.4: (Pre small group) Preference for desalination regressed against candidate preferences.

	Estimate	Std. Error	t value	$Pr(> t)$	
(Intercept)	31.98502	11.90967	2.686	0.00955	**
c1	0.38319	0.11159	3.434	0.00114	**
c2	-0.27543	0.10864	-2.535	0.01411	*
c3	-0.02629	0.08633	-0.305	0.76187	
c4	0.48169	0.10961	4.395	5.11e	05 ***
c5	-0.08414	0.10209	-0.824	0.41342	
FORMULA: desal $\sim$ c1 + c2 + c3 + c4 + c5					
—					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 16.21 on 55 degrees of freedom					
Multiple R-squared: 0.6719, Adjusted R-squared: 0.6421					
F-statistic: 22.52 on 5 and 55 DF, p-value: 3.214e-12					

Table 4.5: (Post small group) Preference for desalination regressed against candidate preferences.

	Estimate	Std. Error	t value	$Pr(> t)$	
(Intercept)	26.995871	23.211431	1.163	0.26742	
c1	0.470627	0.109678	4.291	0.00105	**
c2	-0.329328	0.243230	-1.354	0.20069	
c3	-0.210380	0.122596	-1.716	0.11183	
c4	0.639469	0.177087	3.611	0.00357	**
c5	0.006753	0.171126	0.039	0.96917	
FORMULA: desal $\sim$ c1 + c2 + c3 + c4 + c5					
—					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 11.53 on 12 degrees of freedom					
Multiple R-squared: 0.9294, Adjusted R-squared: 0.9					
F-statistic: 31.62 on 5 and 12 DF, p-value: 1.647e-06					

Table 4.6: Regression after a reflective follow-up discussion, after small and large group events.

candidate opinions – it looks very much like the results from the initial door-to-door survey. On the other hand, Table 4.5 shows the regression just after small group discussion. Attitudes toward conservation and, especially, towards the importance of having a more diverse set of choices for our water supply became statistically significant in predicting the preference for desalination. This relationship was significant in the large door-to-door sample. But the pre small group sample was too small to capture what was very diffuse knowledge about the status and sources of our water supply. This relationship became more broadly relevant to desalination preferences through the process of deliberation, after which the R-squared measure has risen above .6, indicating significant predictive power. From this data we can't tell if this was due to the impact of information provided in the Issue Guide and/or the exchange of ideas and feelings during deliberation, but some clues will emerge in the next section.

The final regression, in Table 4.5, should be considered as only anecdotal evidence, but intriguing. This was a small follow up meeting (18 participants) for individuals who had already been to one or both of the small and large group events. I recruited individuals non-randomly in an attempt to get close to a balance between desalination supporters and opponents. The meeting was reflective in that discussion focused on what was driving the different opinions of pro, anti, and neutral or undecided participants. We tried to clarify what the different needs drove each group and found that the only substantial difference was in the need for information. The pro-desalination group was satisfied that alternative options had been fully explored and felt confident that desalination was the most effective solution. The three hours of discussion in this

group seemed to me to generate the greatest insights into each other's opinions. This was my subjective interpretation of the evening and the R-squared score of .9 provides some additional support.

Decision stability and Condorcet winners

The final section looks at the social choice question directly through an analysis of Condorcet winners in the small groups that deliberated together as well as through a number of virtual comparisons to gain insight into these results. Given our field of five candidates, there are a number of reasonable procedures a group might use to select its choice of preferred candidate. For example, everyone could vote for their top pick and the candidate with the highest number of votes becomes the group's choice. But, with more than two candidates it is possible that one of the other candidates is actually preferred by a majority over the winner, leading to alternative procedures involving runoffs, Borda counts, etc. Moreover, having a majority winner does not eliminate the possibility that there are preference cycles in pairwise comparisons. Transitive group preferences are guaranteed in the case that there is a Condorcet winner.

Candidate X is considered a Condorcet winner if for every binary comparison between candidate X and one of the alternative candidates, Y, candidate X would win the majority of votes over Y. From our preference survey of each small group, we can easily calculate whether there would have been a Condorcet winner. In comparing X and Y, for each member of the group we assign a vote to X if and only if the member

assigned a higher preference level for candidate X than candidate Y. Groups may fail to have a Condorcet winner due to preference cycling, but in small groups with even numbers of members it is also common to have tie votes. Thus, for small groups with random preferences (uniformly distributed from 0 to 100 for each candidate, in our case), the probability of a Condorcet winner is about 75% given an odd number of group members but only 30 to 40% for an even number of group members.

Table 4.7 shows the the results of a Condorcet contest for the preference profiles of each of the small groups, from the pre-meeting survey. The columns show the number of members in the group, which candidate was the Condorcet winner (or "No Winner" in some cases), and the "least margin". Least margin is the smallest difference in votes between the winner and the candidate(s) that came in second place. Clearly, the larger the least margin the more stable the outcome is to small deviations of preferences. At the bottom of the least margin column is the sum of all the winning margins, giving a summary statistic of the Condorcet stability of this 8-group set of preference profiles. Table 4.8 shows how things have changed in the post small group survey.

From the sum of margins we can see that Condorcet stability is significantly greater after deliberation. Given the sum of margins statistic it is now straightforward to evaluate the hypothesis of an impartial culture. We can draw from uniform random preferences to simulate its distribution for this collection of group sizes. The graphics below depict this distribution and the lower image shows points in the distribution where the pre and post small group survey statistics lie.

The first image shows the histogram for all the resulting margin sums for a

simulation with 150,000 samples of random preferences. A large number was used in order to get a reasonable estimate of the tail of the distribution. The second image shows a kernel density estimate for this distribution and the points in the distribution where the pre and post small group surveys lie. These can also be assessed quantitatively by looking at the proportions of runs that are as or more extreme than these values.

For the pre small group survey, the proportion of margin sums that are greater than or equal to 9 is 0.17. This means that we could reject the hypothesis of an impartial culture with 83% confidence. This is indicative but not conclusive.

However, after deliberating in a small group format we can conclusively reject the assumption of impartial culture: the probability of a margin sum value being greater than or equal to 14 is estimated as .0045 from the simulation. We can reject this hypothesis with greater than 99% confidence.

Given that our post-deliberation groups exhibit structure in their preference profiles that is not consistent with impartial culture, we would like to attribute the change to some aspect of the deliberative experience. This could be from reading the Issue Guide and collectively absorbing shared information, the exchange of ideas and comparison of perspectives allowed by the dialog process, or some other aspect. In order to evaluate this I performed another simulation. Let us take the collection of preferences of all small group participants at the pre-group survey as a representative sample, rather than using randomly generated preferences. Indeed, the groups were composed of one potential drawing from this sample distribution and the only thing determining which individuals fell together in a given group was their choice of a meeting time, which is

plausibly unrelated to their preference distribution over desalination candidates.

Our new hypothesis to consider is that deliberation was not the cause of the jump in margin sums, rather it was a random outcome from a new distribution that was not qualitatively different from the incoming pre-group sample. We can evaluate this with a similar approach to the impartial culture simulation, but this time choosing the pre-group discussion sample distribution as representing the population, rather than a uniform random distribution. We then take draws from this population and assign them to groups of the same sizes as in our actual deliberation. Specifically, we sample without replacement to fill a particular group but then reset to the initial sample population before starting the next group. In the end, we calculate the margin sum of the hypothetical groups using the sampled preference profiles.

Results of this simulation can be seen in Figure 4.4, and show that the proba-

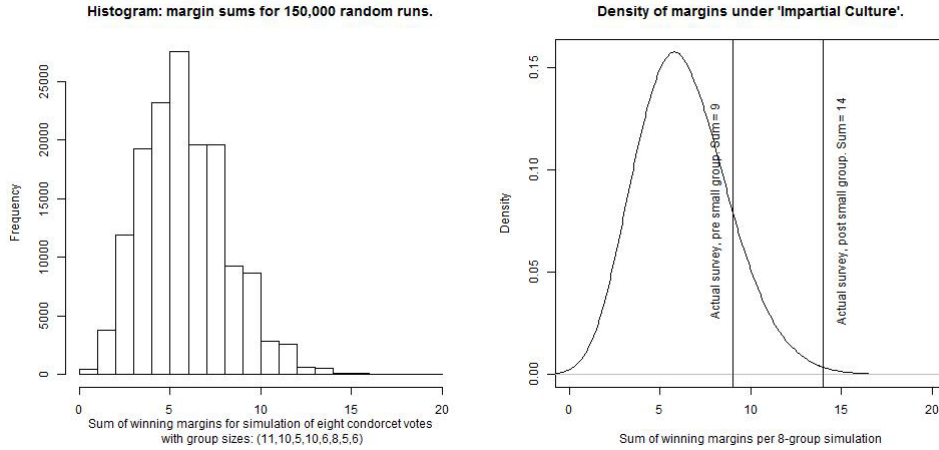


Figure 4.4: Desal Preferences: Left: All door-to-door. Right: The ones who attended

bility of getting a margin sum greater than or equal to 9 is 40%. This is quite reasonable considering we did get a margin sum of 9 in the actual draw. Also, the probability of getting greater than or equal to 14 is only 4.4%, so we reject this with 95% confidence. This is encouraging, but we still don't have a sense of what aspect of the deliberative process led to the increase in preference correlations. Perhaps we could have just emailed the Issue Guide to everyone and had the same effect on preference adjustment?

To evaluate this, we next take the collective set of preferences determined by the post-group survey and consider this as representative of the population distribution. Sampling from this (without replacement within a group, with replacement between groups) reveals that the probability of our statistic being greater than or equal to 14 is actually actually slightly less now, 3.0%. The implication of this, with the caveat that this is one small study, is that most (if not all) of the correlation structure that pushes these groups towards a transitive and more stable Condorcet winner is coming from the dialog with other group members. This is not to say that the information provided in the Issue Guide is irrelevant, indeed they are engaging in dialog about or stimulated by the Issue Guide. The point is that preferences have become correlated within the small groups themselves to increase Condorcet stability, while there is no indication that any broad, out of group, change in decision stability has developed. The face to face interaction through information-based small group deliberation was crucial for overcoming the impossibility result of rational social choice.

If this result can be reliably replicated, it would support the assertions of deliberation enthusiasts that small groups can achieve more coherent and rational decision

Group Size	Winning Candidate #	Least Margin
11	3	4
10	5	1
5	2	1
10	2	1
6	No Winner	0
7	3	1
5	No Winner	0
6	3	1
Sum of margins statistic =		9

Table 4.7: Condorcet results from pre survey for each of the eight small groups.

making through deliberative methods – at least to the extent that intransitive preferences materially contribute to poor group choices. Face to face deliberation was crucial in structuring collective preferences, while studying identical background information failed to show this effect. On the other hand, the fact that this measure was essentially unchanged when sampling from the post-deliberation population casts doubt on the implications of the result for large scale decision making, such as for elected officials or ballot initiatives. Or, this may be a function of the extreme complexity of the question addressed. Regression analysis supported the idea that, globally, preference relationships are becoming increasingly correlated. Water policy is impacted by a overwhelming number of complicated factors. Perhaps two hours of deliberation was insufficient to translate this global increase in cardinal preference correlations into a global rise in ordinal aggregation stability, while still having a significant impact on stability of within group rankings.

Group Size	Winning Candidate #	Least Margin
11	2	2
10	No Winner	0
5	2	3
10	2	3
6	4	2
8	4	2
5	4	2
6	No Winner	0
Sum of margins statistic =		14

Table 4.8: Condorcet results from post survey for each of the eight small groups.

Chapter 5

Conclusion

Bibliography

- [1] Arrow, Kenneth (1953) Social Choice and Individual Values (New York: Wiley).
- [2] Barabas, J. (2004) How Deliberation Affects Policy Opinions, American Political Science Review, 98
- [3] Allahverdyan AE, Galstyan A (2014) Opinion Dynamics with Confirmation Bias. PLoS ONE 9(7): e99557. doi:10.1371/journal.pone.0099557
- [4] Barraclough DJ, Conroy ML, Lee D. Prefrontal cortex and decision making in a mixed-strategy game. Nature Neuroscience. 2004;7:404410.
- [5] Baron, R. S. “So right it’s wrong: Groupthink and the ubiquitous nature of polarized group decision making.” Advances in experimental social psychology 37 (2005): 219-253.
- [6] Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. Journal of Behavioral Decision Making, 16(3), 215-233.
- [7] Barseghyan, L., Molinari, F., O’Donoghue, T., & Teitelbaum, J. C. (2013). The

- nature of risk preferences: Evidence from insurance choices. *American Economic Review*, 103(6), 2499-2529.
- [8] Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 12(3), 151-170.
- [9] Birnbaum, M. H. (1999). Paradoxes of Allais, stochastic dominance, and decision weights. *Decision science and technology: Reflections on the contributions of Ward Edwards*, 27-52.
- [10] Black, Duncan (1948) On the rationale of group decision-making. *The Journal of Political Economy*, 56(1), 23.
- [11] Bleichrodt, H., & Pinto, J. L. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management science*, 46(11), 1485-1496.
- [12] Bohlmann, J. D., Rosa, J. A., Bolton, R. N., & Qualls, W. J. (2006). The effect of group interactions on satisfaction judgments: Satisfaction escalation. *Marketing Science*, 25(4), 301-321.
- [13] Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology*, 36, 1224-1240.
- [14] Camilleri, Adrian R. & Newel, Ben R. (2011). When and why rare events are

- underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychon Bull Rev* (2011) 18:377384, DOI 10.3758/s13423-010-0040-2
- [15] Crano, W. D., & Prislin, R. (2006). Attitudes and persuasion. *Annu. Rev. Psychol.*, 57, 345-374.
- [16] Cronqvist, H., & Siegel, S. (2014). The genetics of investment biases. *Journal of Financial Economics*, 113(2), 215-234.
- [17] Darwin, Charles (1909) "The Origin of Species", P. F. Collier & Son
- [18] Andreea O. Diaconescu, Christoph Mathys, Lilian A. E. Weber, Lars Kasper, Jan Mauer, Klaas E. Stephan; Hierarchical prediction errors in midbrain and septum during social learning, *Social Cognitive and Affective Neuroscience*, Volume 12, Issue 4, 1 April 2017, Pages 618634, <https://doi.org/10.1093/scan/nsw171>
- [19] Dryzek, John S., and Christian List. "Social choice theory and deliberative democracy: a reconciliation." *British Journal of Political Science* 33.1 (2003): 1-28.
- [20] Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C., & Paluck, E. L. (2010). Disaggregating deliberations effects: An experiment within a deliberative poll. *British journal of political science*, 40(2), 333-347.
- [21] Fehr-Duda, H., & Epper, T (2012) Probability and Risk: Foundations and Economic Implications of Probability-Dependent Risk Preferences *Annu. Rev. Econ.* 2012. 4:56793

- [22] Fehr-Duda, H., De Gennaro M, Schubert R (2006) Gender, financial risk, and probability weights. *Theory and Decision* 60: 283-313.H.
- [23] Fishkin, James S. (1991) *Democracy and deliberation: New directions for democratic reform* (Vol. 217). New Haven: Yale University Press.
- [24] Ford Peter, "Why Asia is still fighting over World War II", July 26, 2015, *Christian Science Monitor*. <http://www.csmonitor.com/World/Asia-Pacific/2015/0726/Why-Asia-is-still-fighting-over-World-War-II>
- [25] Fox, C. R., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, 1(2), 159.
- [26] Gaulin, Steven J. C. and Donald H. McBurney. *Evolutionary psychology*. Prentice Hall. 2003.
- [27] Gigone, D., & Hastie, R. (1993). The common knowledge effect: Information sharing and group judgment. *Journal of Personality and Social Psychology*, 65, 959-974.
- [28] Glockner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, 123(1), 21-32.
- [29] Godfrey-Smith, P. (2016). *Other minds: The Octopus, the sea, and the deep origins of consciousness*. Farrar, Straus and Giroux.
- [30] Gonzalez, R., Wu, G. (1999). On the shape of the probability weighting function. *Cognitive psychology*, 38(1), 129-166.

- [31] Grolu, B., Will, G. J., & Crone, E. A. (2014). Neural correlates of advantageous and disadvantageous inequity in sharing decisions. *PloS one*, 9(9), e107996.
- [32] Hansen, K. A., Hillenbrand, S. F., and Ungerleider, L. G. (2012). Human brain activity predicts individual differences in prior knowledge use during decisions. *J. Cogn. Neurosci.* 24, 14621475.
- [33] Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions From Experience and the Effect of Rare Events in Risky Choice. *Psychological Science*, 15, 534539.
- [34] Hertwig, R., & Erev, I. (2009). The descriptionexperience gap in risky choice. *Trends in cognitive sciences*, 13(12), 517-523.
- [35] Hilbig, Benjamin E.; Glockner, Andreas (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition. *Acta Psychologica*, Volume 138, Issue 3, November 2011, Pages 390396
- [36] Ming Hsu, Ian Krajbich, Chen Zhao and Colin F. Camerer (2009) Neural Response to Reward Anticipation under Risk Is Nonlinear in Probabilities. *Journal of Neuroscience* 18 February 2009, 29 (7) 2231-2237; DOI: <https://doi.org/10.1523/JNEUROSCI.5296-08.2009>
- [37] Hytonen, K., Baltussen, G., Van den Assem, M. J., Klucharev, V., Sanfey, A. G., & Smidts, A. (2014). Path dependence in risky choice: Affective and deliberative

- processes in brain and behavior. *Journal of Economic Behavior & Organization*, 107, 566-581.
- [38] Ito, A., Nobuhito Abe, Toshikatsu Fujii, Akiko Hayashi, Aya Ueno, Shunji Mugikura, Shoki Takahashi, Etsuro Moria (2012) The contribution of the dorsolateral prefrontal cortex to the preparation for deception and truth-telling. *Brain Research*, Volume 1464, 29 June 2012, Pages 43-52
- [39] Jacobs, Lawrence R. , Fay Lomax Cook, and Michael X. Delli Carpini. *Talking together: Public deliberation and political participation in America*. University of Chicago Press, 2009.
- [40] Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision*, 50(1), 59-99.
- [41] Kahneman, Daniel and Amos Tversky. (1979). "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* 47, 263-291
- [42] Kahneman, Daniel and Amos Tversky. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5:297-323
- [43] Kameda, T., Tindale, R. S., & Davis, J. H. (2003). 14 Cognitions, Preferences, and Social Sharedness: Past, Present, and Future Directions in Group Decision Making. *Emerging perspectives on judgment and decision research*, 458.
- [44] Kameda, T., Tsukasaki, T., Hastie, R., & Berg, N. (2010, September 6). *Democracy*

- Under Uncertainty: The Wisdom of Crowds and the Free-Rider Problem in Group Decision Making. *Psychological Review*. doi: 10.1037/a0020699
- [45] Kerr, N. L., & Tindale, R. S. (2011). Group-based forecasting?: A social psychological analysis. *International journal of forecasting*, 27(1), 14-40.
- [46] King, A.J., Douglas, C.M.S., Huchard, E., Isaac, N.J.B., and Cowlshaw, G. (2008). Dominance and affiliation mediate despotism in a social primate. *Curr. Biol.* 18, 1833-1838.
- [47] Knoch D, Nitsche MA, Fischbacher U, Eisenegger C, Pascual-Leone A, Fehr E. Studying the Neurobiology of Social Interaction with Transcranial Direct Current Stimulation—The Example of Punishing Unfairness. *Cerebral Cortex*. 2008;18:1987-1990
- [48] Stephen C. Levinson, “Turn-taking in Human Communication Origins and Implications for Language Processing”, *Trends in Cognitive Science* 1514, DOI: 10.1016/j.tics.2015.10.010
- [49] List, C., Luskin, R. C., Fishkin, J. S., & McLean, I. (2013). Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls. *Journal of Politics*, 75(1).
- [50] List, C. (2006). The discursive dilemma and public reason. *Ethics*, 116(2), 362-402.
- [51] Lorenz, J., Rauhut, H., & Kittel, B. (2015). Majoritarian democracy un-

- dermines truth-finding in deliberative committees. *Research & Politics*, 2(2), 2053168015582287.
- [52] Luskin, Robert C. & James S. Fishkin (2002, March) Deliberation and better citizens Paper presented at the Joint Sessions of Workshops of the European Consortium for Political Research , Turin , Italy.
- [53] Maloney, L. T., Wu, S. W., & Dal Martello, M. F. (2006). Movement planning under risk differs from decision making under risk in how subjects make use of probability information. *Journal of Vision*, 6(6), 928-928.
- [54] Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and brain sciences*, 34(02), 57-74.
- [55] Engle, Langer-Osuna and McKinney de Royston (in press) reveal that a students evaluation of evidence is often influenced by their interpretation of the presenter of that evidence rather than an evaluation of its validity.
- [56] Mojzisch, Andreas; Schulz-Hardt, Stefan. Knowing others' preferences degrades the quality of group decisions. *Journal of Personality and Social Psychology* 98.5 (May 2010): 794-808.
- [57] Mojzisch, A., Schulz-Hardt, S., Kerschreiter, R., & Frey, D. (2008). Combined effects of knowledge about others' opinions and anticipation of group discussion on confirmatory information search. *Small Group Research*, 39(2), 203-223.

- [58] Miller, David (1992), *Deliberative Democracy and Social Choice*. *Political Studies*, 40: 5467. doi: 10.1111/j.1467-9248.1992.tb01812.x
- [59] Myers, David G., and Helmut Lamm. "The group polarization phenomenon." *Psychological Bulletin* 83.4 (1976): 602.
- [60] Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology*, 2, 175-220.
- [61] Nunn, P. D.; Reid, N. J. Aboriginal Memories of Inundation of the Australian Coast Dating from More than 7000 Years Ago. *Australian Geographer*, 2015; 1 DOI: 10.1080/00049182.2015.1077539
- [62] Prelec, D. (1998). The probability weighting function. *Econometrica*, 497-527.
- [63] Rabin M, Schrag JL (1999) First Impressions Matter: A Model of Confirmatory Bias. *The Quarterly Journal of Economics* 114: 37. doi: 10.1162/003355399555945
- [64] Rahnev, D., Lau, H., and de Lange, F. P. (2011). Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. *J. Neurosci.* 31, 10741-10748.
- [65] Reimer, T., & Hoffrage, U. (2006). The ecological rationality of simple group heuristics: Effects of group member strategies on decision accuracy. *Theory and Decision*, 60(4), 403-438.
- [66] Richerson, P., Baldini, R., Bell, A., Demps, K., Frost, K., Hillis, V., ... & Zefferman,

- M. (2015). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, 1-71.
- [67] Riker, William H. (1982) *Liberalism Against Populism*. San Francisco: W. H. Freeman.
- [68] Rubin, Paul H. (2003) "Folk Economics." *Southern Economic Journal*, Vol. 70, No. 1, 157-171. DOI: 10.2307/1061637
- [69] Schkade, D., Sunstein, C. R., & Kahneman, D. (2000). Deliberating about dollars: The severity shift. *Columbia Law Review*, 1139-1175.
- [70] Juliana Schroeder, Michael Kardas, Nicholas Epley. The Humanizing Voice: Speech Reveals, and Text Conceals, a More Thoughtful Mind in the Midst of Disagreement. *Psychological Science*, 2017; 095679761771379 DOI: 10.1177/0956797617713798
- [71] Schulz-Hardt, Stefan; Frey, Dieter; Lthgens, Carsten; Moscovici, Serge (2000). "Biased information search in group decision making". *Journal of Personality and Social Psychology*, Vol 78(4), Apr 2000, 655-669. <http://dx.doi.org/10.1037/0022-3514.78.4.655>
- [72] Schnuerch, R., & Gibbons, H. (2015). Social proof in the human brain: Electrophysiological signatures of agreement and disagreement with the majority. *Psychophysiology*.
- [73] Sloman, S., & Fernbach, P. (2017). *The knowledge illusion: why we never think alone*. Penguin.

- [74] Smith, A. (1784). *An Inquiry Into the Nature and Causes of the Wealth of Nations*:
By Adam Smith... W. Strahan.
- [75] Stasser, G., Kerr, N. L., & Davis, J. H. (1989). Influence processes and consensus
models in decision-making groups, in P. Paulus (Ed.), *Psychology of group influence*
(2nd ed.). Hillsdale, N J: Erlbaum.
- [76] Stauffer, W. R., Lak, A., Bossaerts, P., & Schultz, W. (2015). Economic choices
reveal probability distortion in macaque monkeys. *The Journal of Neuroscience*,
35(7), 3146-3154.
- [77] Stoner, J.A. (1961). "A comparison of individual and group decision involving risk".
Unpublished master's thesis, Massachusetts Institute of Technology.
- [78] Sunstein, Cass R. "The law of group polarization." *Journal of Political Philosophy*
10.2 (2002): 175-195.
- [79] Suzuki, S., Jensen, E. L., Bossaerts, P., & O'Doherty, J. P. (2016). Behavioral
contagion during learning about another agents risk-preferences acts on the neural
representation of decision-risk. *Proceedings of the National Academy of Sciences*,
113(14), 3755-3760.
- [80] Tobler, P. N., Christopoulos, G. I., O'Doherty, J. P., Dolan, R. J., & Schultz, W.
(2008). Neuronal distortions of reward probability without choice. *The Journal of*
Neuroscience, 28(45), 11703-11711.
- [81] Tymula, A., Belmaker, L. A. R., Ruderman, L., Glimcher, P. W., & Levy, I. (2013).

- Like cognitive function, decision making across the life span shows profound age-related changes. *Proceedings of the National Academy of Sciences*, 110(42), 17143-17148.
- [82] Ungemach C, Chater N, Stewart N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychol Sci.* 2009 Apr;20(4):473-9.
- [83] Wouter van den Bos, Eric van Dijk, Michiel Westenberg, Serge A.R.B. Rombouts, Eveline A. Crone. Changing Brains, Changing Perspectives: The Neurocognitive Development of Reciprocity. *Psychological Science* 22(1) 6070. DOI: 10.1177/0956797610391102
- [84] Weibull, J. W. (1997). *Evolutionary game theory*. MIT press.
- [85] West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of personality and social psychology*, 103(3), 506.
- [86] Wu, S. W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences*, 106(15), 6088-6093.
- [87] Wu, S. W., Delgado, M. R., & Maloney, L. T. (2011). The neural correlates of subjective utility of monetary outcome and probability weight in economic and in motor decision under risk. *The Journal of Neuroscience*, 31(24), 8822-8831.

- [88] Yoshida W, Seymour B, Friston KJ, Dolan RJ. Neural mechanisms of belief inference during cooperative games, *Journal of Neuroscience* , 2010, vol. 30 32 (pg. 10744-51)
- [89] Zhong, S., Israel, S., Xue, H., Ebstein, R. P., & Chew, S. H. (2009). Monoamine oxidase A gene (MAOA) associated with attitude towards longshot risks. *PLoS One*, 4(12), e8516.
- [90] Zhong, S., Chew, S. H., Set, E., Zhang, J., Xue, H., Sham, P. C., ... & Israel, S. (2009). The heritability of attitude toward economic risk. *Twin Research and Human Genetics*, 12(1), 103-107.