

# **Debating Deliberative Democracy: How Deliberation Changes the Way People Reason**

Thomas William Flynn

A thesis submitted in the fulfilment of the requirements of the degree of Doctor of  
Philosophy (PhD)

University of York: Department of Politics

January 2011

# ABSTRACT

---

The concepts of deliberation and deliberative democracy have attracted much attention in political theory over the past twenty years. At first seen as both highly idealised and unreflective of reality, they have now shed this accusation of impracticality, as practitioners and policy makers alike have attempted to institute deliberative principles on a national and international scale. Running alongside this has been the desire to both understand political deliberation and its effects more fully, and to then apply this new information back to deliberative democratic theory. This thesis sits in the latter tradition, presenting an empirical investigation of political deliberation and then discussing how it relates back to deliberative models of democracy. Where it departs from all of the contemporary experimental work, however, is the methodology and conceptual model it is founded upon. Embracing the decision and game theoretic approaches, I develop a three-fold framework to study the effects of deliberation on individual decision-making. After outlining two levels of ‘preference’ and ‘issue’, I focus on the third, which I term *agency*. I then compare a particular case of agency revision, which moves people from individualistic to *team reasoning*, before developing and putting into action an experimental test of the phenomenon. Finally, I then combine these results with the most recent drive in deliberative democracy towards a systemic approach, and derive an alternative, more positive argument for this recasting.

# CONTENTS

---

I.	List of Tables	p. 5
II.	List of Figures	p. 6
III.	Acknowledgements	p. 7
IV.	Author's Declaration	p. 8
V.	Introduction	p. 9
<b>1</b>	<b>Deliberation and the Deliberative Turn in Democratic Theory</b>	<b>p. 11</b>
1.1	Introduction	p. 11
1.2	Defining Deliberative Democratic Theory	p. 11
1.3	Defining Deliberation	p. 25
1.4	Concluding Remarks	p. 41
<b>2</b>	<b>A Rational Choice Approach to Deliberative Revision: Preference, Issue and Agency</b>	<b>p. 42</b>
2.1	Introduction	p. 42
2.2	Deliberation and Preference Revision	p. 43
2.3	Deliberation and Issue Revision	p. 57
2.4	Deliberation and Agency Revision	p. 59
2.5	Concluding Remarks	p. 72
<b>3</b>	<b>Framing and Team Reasoning: How Deliberation Primes Social Identity</b>	<b>p. 73</b>
3.1	Introduction	p. 73
3.2	Identity Theory and Rational Choice: Defining the Self	p. 74
3.3	Deliberation and Social Identity	p. 86
3.4	Concluding Remarks	p. 94
<b>4</b>	<b>Measuring Team Reasoning: An Experimental Economics Methodology</b>	<b>p. 95</b>
4.1	Introduction	p. 95
4.2	Why Experimental Economics?	p. 95
4.3	Problems with an Experimental Economics Approach	p. 106
4.4	Constructing Game Theoretic Encounters	p. 112
4.5	Concluding Remarks	p. 118
<b>5</b>	<b>The Case Study of Deliberation: A People's Inquiry into Personal and Private Information</b>	<b>p. 120</b>
5.1	Introduction	p. 120
5.2	Deliberation in Action	p. 121
5.3	Empirical Criteria for High Quality Deliberation	p. 125
5.4	The Case Study: Personal and Private Information	p. 132
5.5	The Case Study: How Does it Compare?	p. 137
5.6	Concluding Remarks	p. 140
<b>6</b>	<b>Results and Analysis: Does Deliberation Trigger Agency Revision?</b>	<b>p. 141</b>
6.1	Introduction	p. 141

6.2	Statistical Methodology	p. 142
6.3	Results and Statistical Analysis	p. 144
6.4	Further Analysis	p. 152
6.5	Discussion and Concluding Remarks	p. 165
<b>7</b>	<b>Conclusion: Recasting Deliberation and Deliberative Democracy</b>	<b>p. 168</b>
7.1	Introduction	p. 168
7.2	Third Generation Deliberative Democracy	p. 169
7.3	Agency Revision and Deliberative Democratic Systems	p. 186
7.4	Concluding Remarks and Further Research	p. 192
VI	Appendix 1: The Experiment	p. 193
VII	Bibliography and References	p. 197

# I. LIST OF TABLES

---

Table 2.1	Individual preferences over two alternative pre deliberation	p. 50
Table 2.2	Individual preferences over two alternatives post deliberation	p. 51
Table 2.3	Individual preferences over three alternatives pre deliberation	p. 54
Table 2.4	Individual preferences over three alternatives post deliberation	p. 54
Table 2.5	‘Bundled’ preferences over alternatives pre deliberation	p. 58
Table 2.6	‘Unbundled’ preferences over alternatives post deliberation	p. 59
Table 5.1	Sex of Participants	p. 133
Table 5.2	Age of Participants	p. 133
Table 5.3	Ethnicity of Participants	p. 134
Table 5.4	NRS Social Grade of Participants	p. 134
Table 6.1	Experimental Results for Game One	p. 144
Table 6.2	Experimental Results for Game Two	p. 145
Table 6.3	Experimental Results for Game Three	p. 145
Table 6.4	Logistic Regression Parameter Estimates for Game One	p. 147
Table 6.5	Logistic Regression Parameter Estimates for Game Two	p. 147
Table 6.6	Logistic Regression Parameter Estimates for Game Three	p. 148
Table 6.7	Experimental Results for Game Four	p. 149
Table 6.8	Logistic Regression Parameter Estimates for Game Four	p. 150
Table 6.9	Experimental Results for Game Five	p. 151
Table 6.10	Logistic Regression Parameter Estimates for Game Five	p. 151
Table 7.1	Features of Different Models of Macro Deliberation	p. 179

## II. LIST OF FIGURES

---

Fig 1.1	A Taxonomy of Democratic Theories	p. 14
Fig 1.2	Elements of Preference Construction	p. 23
Fig 1.3	Three Levels of Deliberative Criteria	p. 27
Fig 1.4	The Behavioural Account of Deliberation	p. 30
Fig 1.5	The Structural Account of Deliberation	p. 39
Fig 2.1	Preferences pre deliberation	p. 51
Fig 2.2	Preferences post deliberation	p. 52
Fig 2.3	The Hi Lo Game	p. 64
Fig 2.4	A Prisoners' Dilemma Game	p. 68
Fig 2.5	The Enlightenment Trail	p. 69
Fig 3.1	The Continuum of Identity	p. 77
Fig 4.1	The Methodology of Experimental Economics	p. 96
Fig 4.2	The Stag Hunt	p. 101
Fig 4.3	A 3x3 Game Theoretic Interaction	p. 102
Fig 4.4	A Decomposable Prisoners' Dilemma Game	p. 104
Fig 4.5	A Non Decomposable Prisoners' Dilemma Game	p. 106
Fig 4.6	Game Theoretic Encounter One	p. 113
Fig 4.7	Game Theoretic Encounter Two	p. 114
Fig 4.8	Game Theoretic Encounter Three	p. 114
Fig 4.9	Game Theoretic Encounter Four	p. 115
Fig 4.10	Game Theoretic Encounter Five	p. 116
Fig 6.1	Team Reason Strategy Choices in Games One to Three	p. 146
Fig 6.2	Team Reason Strategy Choices in Game Four	p. 149
Fig 6.3	Team Reason Strategy Choices in Game Five	p. 151
Fig 7.1	Parallel Deliberation	p. 171
Fig 7.2	Serial Deliberation	p. 173
Fig 7.3	Disjointed Deliberation	p. 174
Fig 7.4	'Parallel' Distributed Deliberation	p. 176
Fig 7.5	'Disjointed' Distributed Deliberation	p. 177

### III. ACKNOWLEDGEMENTS

---

Without a doubt, the four years that I spent in the Department of Politics at the University of York have been both the most stimulating and enjoyable period of my life. I have had the opportunity to meet some wonderful people, and for that I am truly grateful.

More than anyone involved in my work, I want to thank my supervisor Dr John Parkinson for his unwavering support over the course of my studies. This thesis would not have been possible without his constant enthusiasm, patience and advice. I would also like to say a particular thank you to Lisa Webster in the department, who often moved heaven and earth to accommodate my teaching commitments. Others that have given me time and expertise include John Bone, Arthur Kang'ombe, John Dryzek, Selen Ayirtman, Simon Niemeyer, Keith Dowding, Andrew Colman, Heather Savigny, and Robert Sugden.

A special thank you also goes out to the staff of Wentworth Graduate College, who made University halls of residence into a home for nearly three years. To Carl and Karen Thompson, Derek Lavin, David Hammond, Cath Henderson and Angela Grayson, I will miss you all.

A number of close friends were also crucial to completing this project, and helped motivate me when work seemed too much. To Charlie Leyland, Lewis Bretts, Matt Burton, Jamie Tyler, Nick Scarlett and Anne-Marie Canning, without you all, I know I couldn't have survived.

Finally, I want to say thank you to my family. Without their support and self-sacrifice, I would never have got this far in my studies. I owe them everything, and dedicate this thesis to my mother Gillian, my father Stephen, and my two brothers Alex and Neil.

## **IV. AUTHOR'S DECLARATION**

---

I certify that the thesis presented by me in 2011 for the degree of PhD in the Department of Politics at the University of York, is solely my own work other than where I have clearly indicated that it is the work of others, and that the extent of any work carried out jointly by myself and any other person is clearly identified in it.



## V. INTRODUCTION

---

In a 1998 survey article, James Bohman famously lamented the lack of any comprehensive empirical research programme undertaken in the field of deliberative democracy. In particular, he threw the spotlight onto the crucial link between empirical evidence and institutional design, and argued this relationship was vital if deliberative democratic theory was to make the difficult transition from political theory to political reality:

For all the sophistication of these discussions of justification, institutional design and feasibility, there is still a surprising lack of empirical case studies of democratic deliberation at the appropriate level and scale.

(Bohman 1998; p. 419)

This relationship between theory and practice, he went on to argue, was critical because it helped political scientists gain further insight into the approach from both perspectives. Studying how deliberation happened in the real world and what it actually achieved was crucial for both putting it into practice, as well as then re evaluating the model upon which it was based. In short, whilst as an area in idealised political theory it had indeed ‘come of age’; this had yet to be matched with any equivalent progress in political science. Driven by this assertion that ‘the deliberative model of democracy [had begun to exemplify the] widening gap between normative and empirical approaches to politics’ (Habermas 2006), a significant amount of research has begun to take place on exactly this front (Thompson 2008). In fact, this agenda has developed so much over the past five to ten years, that Dryzek (2008) recently claimed we had witnessed the ‘empirical turn in deliberative democracy’.

With this context in mind, this thesis is a conceptual and empirical investigation of deliberation and through that, deliberative democratic theory. But as I will show in chapter two, the current work taking place in one particular subfield has failed to really grasp the full story behind a fundamental claim made by all deliberative democrats: that the process causes individuals to ‘change’ over the course of the deliberation. The argument structure of this thesis can therefore be summarised as follows:

Chapter	Main Concern
One	To provide a literature review of deliberation and deliberative democracy, and show how both have already been through recasting on the basis of empirical evidence and analysis.
Two	To focus specifically on deliberation, and outline an analytical model to investigate the central claim of deliberative revision.
Three	To further theoretically investigate one particular claim regarding deliberative revision: that it causes individuals to change the way they reason (to team-reasoning) during decision-making.
Four	To outline an experimental methodology used to test empirically for the process of agency revision to team reasoning during political deliberation.
Five	To outline the case study of the mini public used for the experimental investigation of deliberation.
Six	To describe the results from the deliberative revision experiment, and to analyse the various relationships found within the data.
Seven	To then re apply the results and conceptual lessons back to deliberation, and thus deliberative models of democracy.

# — CHAPTER 1 —

## DELIBERATION AND THE DELIBERATIVE TURN IN DELIBERATIVE DEMOCRACY

### 1.1 INTRODUCTION

---

As I stated in the introduction, this thesis is a conceptual and empirical investigation of deliberation, and through this, deliberative democracy. As such, my first task is to provide a discussion of the concepts and debates that have characterised the subject area. This first chapter, then, has three objectives:

- i. To provide a clear, and much needed distinction between the concepts of ‘deliberation’ and ‘deliberative democracy’.<sup>1</sup>
- ii. To provide an overview of deliberative democratic theory, exploring the various challenges to it, and the way the theory has already been recast a number of times due to critical analysis and evidence.
- iii. To explore exactly what is meant by the concept of ‘deliberation’ and to identify the main principles that have been associated with its use in deliberative democratic theory. As I will argue a little later, this is essential for a study with any conceptual or empirical element in its investigation.

In accomplishing these goals, I will demonstrate the degree to which deliberation and deliberative democratic theory have already been through significant reformulation on the basis of critical analysis and evidence. This final point sets the precedent for the conceptual examination in chapters two and three, the resulting empirical investigation based on these lessons during chapters four, five, and six, and the argument for further recasting of deliberation and deliberative democracy that I make in chapter seven.

### 1.2 DEFINING DELIBERATIVE DEMOCRATIC THEORY

---

<sup>1</sup> This point is both crucial for the thesis, as well as the literature at large. A number of recent important review articles of these issues, including Bächtiger et al (2010) and Mansbridge et al (2010), seem to switch too readily between the two concepts. My objective is to clearly distinguish between them during this chapter.

Over the past fifteen years, the debate within the discipline of democratic theory has been dominated by two distinct approaches. Firstly, there is the aggregative approach, which considers raw preferences as the primary material for decision-making, and conceptualises democracy as an economic market where bargaining and self interest reign supreme. Stemming from a highly liberal position, it stipulates that personal preferences need no public or private justification, and aside from their roles in prediction or correction, it regards reasons as insignificant and inconsequential to the process. Collective political decisions are then made according to a specific, although not universal, aggregation method. If a particular threshold is met, usually some form of majoritarian result, then the outcome is binding and deemed legitimate for all citizens.

This model dominated the theoretical literature up until the late 1980s, and arguably still occupies centre stage in the more applied field of democratisation. Its pedigree can be seen in the sheer volume of literature premised on its fundamental principles. Wollheim (1962; p.76), for example, envisaged the democratic process as a type of machine, into which are 'fed, at fixed intervals, the choices of individual citizens'; Riker (1961) sees it as the mere 'summing of preferences', whilst Sen (1970; pp.35-36) describes it as a 'collective choice rule' used to transpose individual preferences into a unique social ordering of alternatives. Mansbridge on the other hand, offers the following comprehensive definition:

Voters pursue their individual interests by making demands on the political system in proportion to the intensity of their feelings. Politicians, also pursuing their own interests, adopt policies that buy them votes, thus ensuring accountability. In order to stay in office, politicians act like entrepreneurs and brokers, looking for formulations that satisfy as many, and alienate as few, interests as possible. From the interchange between self-interested voters and self-interested brokers emerge decisions that come as close as possible to a balanced aggregation of individual interests.

(Mansbridge 1980; p. 17)

The key point to take is that this model of democracy sees only the decision rule as the source of authority, settling what collective decision should be taken, and therefore making the minority of citizens who might disagree, obey it. Since the late 1980s, however, democratic theory has experienced what Dryzek (2000, p. v) famously calls

the 'deliberative turn'<sup>2</sup>, where the focus of democracy has shifted to what Parkinson (2006a; p.1) claims can roughly be described as 'a way of thinking about politics which emphasises the give and take of public reasoning between citizens, rather than counting the votes or authority of representatives'. Moreover, the deliberative conception of democracy stresses the importance of the process itself, whereby 'individuals are amenable to changing their judgements, preferences and views during the course of their interactions, which involve persuasion rather than coercion, manipulation or deception' (Dryzek 2000; p.1). Put more precisely, the focus of democratic theory has shifted starkly from the 'what' question of decision making, to the 'why'.

### 1.2.1 CLASSIFYING MODELS OF DELIBERATIVE DEMOCRACY

Deliberative democratic theory then, formulated largely as a response to the rather more minimal account of democracy espoused above, has a rich theoretical history. This section will draw on a useful distinction made in a number of places including Elstub (2010), Bächtiger et al (2010) and Mansbridge et al (2010), whereby three distinct 'generations' of deliberative democracy can be identified. The first generation<sup>3</sup> of deliberative democrats, including Habermas (1987, 1990, 1994, 1996a, 1996b, 1997), Rawls (1993, 1997a, 1997b), and Cohen (1997), although differing in terms of their focus, all envisaged the process of deliberation as a highly idealised method that resulted a 'superior' collective decision (often resulting in consensus). Second-generation<sup>4</sup> deliberative democrats on the other hand, including Dryzek (1990, 2000), Young (1996, 1999) Goodin (2003), and Gutmann and Thompson (1996, 2004), have attempted to adapt deliberative models to take into account the effects of deep disagreement, other forms of communication and private preferences – therefore rejecting the strict outcome requirement of consensus. Each model has taken a slightly different focus, and thus they are in fact more a collection of interdependent departures from the idealised accounts of their predecessors, than a unified set of approaches. Finally, third generation deliberative democrats<sup>5</sup> are categorised by their desire to explore the ways in which these second generation models might be institutionalised in large modern societies, and can broadly be split into two tracts. One, which has

---

<sup>2</sup> The phrase 'deliberative democracy' was coined originally by Bessette (1980) to describe the discussions between members of the US House of Congress.

<sup>3</sup> Bächtiger et al (2010) label this 'type I', whilst Mansbridge et al (2010) use the term 'classic deliberation' to represent the same concept.

<sup>4</sup> Bächtiger et al (2010) term this 'type II', whilst Mansbridge et al (2010; p. 67) describe it as an '*expansion of the classic ideal*'.

<sup>5</sup> Only Elstub (2010) makes this important distinction clear, although it is certainly implicit in Bächtiger et al (2010).

attempted to show how various micro models might be adapted to mass publics (Ackerman and Fishkin 2002, 2005; Fishkin 1995), and another that has sought to recast deliberative principles in terms of a macro scale ‘system’ (Hendriks 2006; Mansbridge 1999; Dryzek 2009; Goodin 2005; Parkinson 2006a, 2008a). A discussion of these third generation models is taken up in the concluding chapter of this thesis, which takes into account the conceptual and empirical lessons that I draw from the analysis of second-generation deliberation in the following investigative chapters.

Before I look at the movement from first to second-generation incarnations then, I want to briefly address a popular oversimplification/misinterpretation of the way these models can be classified. A number of authors have sought to delineate between theories of deliberative democracy on the grounds of ‘preference formation’ and ‘decision rule’ alone (Fishkin 2005; Shapiro 2003):

Mode of Preference Formation	Decision Rule is Consensus	Decision Rule is Aggregative
None	N/A <sup>6</sup>	Purely Aggregative
Deliberation	First Generation/ Ideal Deliberative	Second Generation/ Deliberative Aggregation

Figure 1.1: A Taxonomy of Democratic Theories

But critically, this fails to fully comprehend the degree to which second generation models differ from their first generation predecessors, and further still, the degree to which second generation models differ from each other. Moreover, it also places far too much emphasis on applying aggregative principles to deliberative models. There is much more that separates these two types of approaches than simply the collective decision of the deliberating individuals – a nuance that will become clear in section 1.2.3. In the next two sections of this chapter then, by outlining first and second-generation accounts of deliberative democracy, the objective is to show how the approach has already been through significant recasting on the basis of contestation and evidence.

<sup>6</sup> It is questionable over whether a) this box exists (how do we know we are in consensus if no deliberation or aggregation can take place), or indeed b) whether it can be construed in democratic terms (it might be seen as representing a totalitarian regime of government).

## 1.2.2 FIRST GENERATION MODELS

Historically of course, the idea that democracy must include deliberation is nothing new, dating back to amongst others, the work of Aristotle (1998; 1287a34) who famously proclaimed the law 'reason unaffected by desire'. In terms of outlining a cohesive theory that models the principles, methods and outcomes of a deliberative democracy however, two authors more than any other are central. One is John Rawls (1993, 1997a, 1997b), most effectively applied in Cohen (1997), who is famous for framing deliberative democracy in terms of a 'moral requirement' (Freeman 2000; p.379). The other is Jurgen Habermas, who proceeds on the basis of discourse ethics and rationality. Both types of account discuss how collective decisions are conceived and endorse consensus as a possible (Cohen) or even required (Rawls/Habermas) ideal. For example Cohen (1997; p.75) states 'ideal deliberation aims to arrive at a rationally motivated consensus', whilst Habermas (1996; p.110) asserts 'the democratic principle states that only those statutes may claim legitimacy that can meet with the assent of all citizens'. Further still, there is also broad agreement between the two perspectives on the nature of the collective outcomes that deliberative democracy will yield. Cohen (1997; p.67) claims that public deliberation 'shapes the identity and interests of citizens in ways that contribute to the formation of a public conception of common good', with Habermas, albeit placing more emphasis on the required procedural conditions, offering a similar statement:

Under the pragmatic presuppositions of an inclusive and non-coercive rational discourse among free and equal participants, everyone is required to take the perspective of everyone else, and thus project herself into the understandings of self and world of all others; from this interlocking of perspectives there emerges an ideally extended we-perspective from which all can test in common whether they wish to make a controversial norm the basis of their shared practice [...]

(Habermas 1995; pp.117-118)

Where they differ, is in the mechanism they argue makes this possible, and the scope to which these principles should be applied. Rawls and Cohen appeal to a strictly defined concept of public reason, applicable only in the public political forum to questions of 'constitutional essentials' (Rawls 2001; p. 41), and which, in a democratic citizenry, is described as the 'reason of its citizens, of those sharing the status of equal citizenship', with its content centred on 'the good of the public' (Rawls 1993; p.213). Consequently

this notion then extends to an interpretation of consensus as that of an 'overlapping consensus', whereby substantive goals are agreed to by people from all moral doctrines from their respectively different personal grounds.

Habermas, on the other hand, relies on both a different interpretation of what deliberation entails (i.e. issues of public reason), as well as the domain in which it is applicable. For him, the process of deliberation should not be restricted in either of the manners characterised by Rawls: it should include appeals to individual moral considerations, and be associated with unofficial networks of private people beyond that of the official political domain (indeed Habermas sees these unofficial domains as the true source of legitimacy for a government). His approach is further grounded on the presence of a number of conditions required for an ideal speech situation, requiring individuals to deliberate with reference to an 'ideal audience or an ideally inclusive community' (Habermas 1996; p. 322), and to ascertain the position that such a community would agree to, under hypothetical conditions of perfectly democratic discourse. This mechanism, referred to as a process of communicative rationality, is indeed often compared to Rawls, but the key point of departure is that Habermas claims without actual deliberation amongst equals, in a situation that approximates his given conditions, then no consensus can be attained.<sup>7</sup>

Each of these two perspectives, then, relies on different primary arguments to defend their respective accounts of deliberative democracy. Rawls and Cohen appeal explicitly to the idea that decisions made will reflect justice and fairness, whilst Habermas makes legitimacy and rationality his core concerns. Each claims that the true site of deliberation in a democracy is different. Rawls and Cohen favour formal political institutions, whilst Habermas puts much more emphasis on the informal public sphere. But where they find common ground, and why they are considered under the same banner, is the rather idealised way in which they model the deliberation in deliberative democracy. Whether made in terms of 'public reason', or 'communicative rationality', both perspectives present a highly normative framework that offers a very strict normative picture of what deliberation should involve. It is principally this feature, not

---

<sup>7</sup> Indeed this very reason is often given for an argument that Rawls' original work does not class him as a deliberative democrat (Saward 2002), in contrast to Habermas whose emphasis on the essentiality of a dialogical process clearly lends itself to the notion of deliberation. This point will become even more relevant in section three of this chapter.



only the fact that they aim for consensus<sup>8</sup>, which ties these approaches together. It is also this feature that I'll return to in section three of this chapter.

### 1.2.3 SECOND GENERATION MODELS

Second-generation models, I want to emphasise once again, are different from their predecessors on a number of different dimensions. In this section, I discuss a selection of these theories by way of their main departing points from first generation ideals. But crucially, I am going to partition off any discussions of how these adaptations might impact on how we define deliberation (rather than deliberative democracy). I do this for reasons of clarity of argument and simplicity. This project is foremost an empirical investigation of deliberation, with the *results* then re applied to deliberative democratic theory. As such, keeping the two (related) concepts distinct is imperative.

#### *Exclusivity Challenges to Deliberative Democracy*

Splitting these second generation models up into different families, I want to consider particular types of objections to (and thus reformulations of) deliberative democratic theory in turn. The first are a collection that might be labelled as inclusivity challenges (Young 1999), i.e. what is included in a deliberative model of democracy. I now consider a number of these in turn.

#### i. What kinds of participation are included in a deliberative democracy?

Within this sub-categorisation, the first departure stems from an argument that Young (2001) labels 'the activist challenge'. She begins the article by pointing out that contemporary democratic theory does little to value, or even include, the role that demonstration or direct action plays in a well functioning democracy. This, it is argued, is problematic, as a number of rights that are considered beneficial to individuals have been secured in exactly this manner. Her examples cite 'the eight hour day', 'votes for women' and 'the right to sit at any lunch counter' – but it is easily conceivable to think of numerous other instances throughout history that would fall into this category. The peaceful protests and civil disobedience pioneered by Ghandi during India's fight for independence, the famous refusal of individuals to pay Thatcher's Poll Tax, or the well-publicised music concerts and rallies that were organised to put pressure on the G8

---

<sup>8</sup> See Manin (1987) for a discussion of how first generation models of deliberative democracy do not necessarily lead to consensus.

nations to write off third world debt. All show that direct political action, namely in the shape of activism, can have a significant positive impact on democracy (Platt 2008).

Young then goes on to offer two types of argument as to why deliberative democratic principles offer negative judgements on behaviour that is arguably essential for society. Both rely on an apparent contradiction between activist behaviour and the norms of deliberative democracy articulated in first generation models. Activist behaviour is categorised as (i) 'interest based', and (ii) unreasonable in nature. The former refers to the point that activists are defined by the very fact that they have taken a particular stance on a particular issue. Emmeline Pankhurst, for example, did not enter the political stage as an individual uncommitted on the issue – she was firmly on the side of women's suffrage from the off. The latter point of unreasonableness, on the other hand, relates to the unwillingness of activists to consider other points of view during the political process. As Young (2001; p. 674) puts it herself: '[activists] simply aim to win the most for their group and engage in power politics to do so'. These characteristics appear quite antithetical to the maxims of public reason and communicative rationality. As the argument logically proceeds then, activists should therefore be excluded from the deliberative decision-making process.

So how does Young propose this problem is resolved? One way is by demonstrating that activist behaviour is actually compatible with the first generation normative ideals. Being 'interest based' is not the same as being 'self-interest based', as it is directed towards an objective that is both principled and communicative in nature. It is also relevant, Young claims, that most activists are often much more likely to have 'good reasons for what they do'. This is indeed a reformulative approach then, but not in the sense of reformulating deliberative democratic theory. Rather, the approach runs in the other direction as it attempts to recast activism in terms of deliberative principles. An alternative way, and one that Young concludes with, is to adapt the deliberative model of democracy to deal with this challenge:

We can conceive the exchange of ideas and processes of communication taking place in a vibrant democracy as far more rowdy, disorderly, and decentred [...] in this alternative conceptualisation, processes of engaged and responsible democratic communication include street demonstrations and sit ins, musical works and cartoons...

(Young 2001; p.688)

By allowing activist forms of participation such as protests into the process, Young effectively shuts down this objection to/critical observation of, deliberative models of democracy. Whilst any model must include reasoned political discussion, there is nothing to stop it from including other forms of participation as well. In doing so, the ‘activist challenge’ no longer applies.

ii. Who participates in a deliberative democracy?

A second important point can be made about participation here. First generation theorists view the deliberative process very much through the lens of an extremely large discursive forum. All individuals who are affected by the collective decision are expected to participate, and moreover, if the ‘ideal’ is to be realised, then this must all happen simultaneously. However, as Dahl’s ‘back of the envelope’ calculation demonstrates, there is a significant problem of scale that deliberative democracy needs to confront:

If an association were to make one decision a day, allow ten hours a day for discussion, and permit each member just ten minutes – rather extreme assumptions [...] – then the association could not have more than sixty members.

(Dahl 1970; pp. 67-68)

A number of non-mutually exclusive ways have popularly been used to reformulate deliberative democracy in light of this objection on grounds of legitimacy (Dryzek 2001, Parkinson 2006a). The first suggests that deliberative models of democracy are only applied to a very small set of possible collective decisions that need to be made – mirroring Rawls’ argument over constitutional essentials. But it should be obvious that this does not really address the problem of scale, as even a single constitutional problem would still take years to decide in a polity of thousands, let alone millions. The second solution, based on Goodin (2000, 2003) and Goodin and Niemeyer (2003), offers an argument for the prioritisation of ‘internal reflection’ over deliberation, meaning individuals are no longer required to physically interact with each other. This is something that I will cover specifically in section three of this chapter when I consider whether deliberation must be a form of external communication.

Dealing with a third possible solution then. Both Rawlsian and Habermasian models of deliberative democracy are highly participatory in nature. But we need look no further than wholly aggregative accounts, which only require individuals to tick boxes on ballot papers, to find that some element of representation is necessary for it to work on a large scale. As a consequence, Bessette (1994) Gastil (2000) amongst others, explicitly envisage a deliberative democracy as one where deliberation plays a role in the election of representatives, although as Dryzek (2001) points out, it is not exactly clear how the problem of scale does not then simply apply to the election campaign itself. Parkinson (2003, 2006b, 2007) on the other hand, offers a solution that approaches the problem from the other perspective. Implicit in his defence of the British House of Lords (which are appointed, not elected), is the idea that an elected representative chamber itself might be conceptualised as the site of deliberative democracy. In this approach which Goodin (2000) labels ‘ersatz deliberation’, individuals only need to elect a limited selection of paid representatives, who then clearly have the both the time and resources to commit themselves fully to the process. However, this in turn brings up a whole set of different legitimacy problems. Should individuals elect representatives that most strongly represent their initial preferences? Or should they elect people who most strongly represent their values, in an attempt to ‘second guess’ their deliberative judgements? (Saward 2006). These questions are crucial for deliberative democrats who wish to reformulate deliberative democracy on these grounds<sup>9</sup>, and indeed, are some of the key reasons behind Dryzek’s (2001, with Niemeyer 2008) recasting in terms of ‘discursive’ rather than deliberative representation.<sup>10</sup>

The problem of scale, clearly, is extremely difficult to overcome. Indeed neither of these possible solutions seem to provide a reformulation that clearly addresses all the specific issues that large-scale deliberative democracy produces. An alternative, however, is suggested by third-generation deliberative democrats, who following Habermas’ (1996) ‘two track model’, have suggested deliberative democracy needs to be recast in more ‘macro’ terms. This represents the most recent shift in the theory, and is something I will consider in depth, along side my experimental results, during the concluding seventh chapter of the thesis.

---

<sup>9</sup> Parkinson (2003) brings in Catt’s (1999) distinction between the ‘delegate’ and ‘trustee’ model of representation to make this point even clearer.

<sup>10</sup> Discursive, rather than deliberative, representation attempts to separate legitimate decision-making from head counts. In this sense, the democratic system is modelled as the confrontation of ‘discourses’ rather than individuals. Parkinson (2003; p.186) offers a nice critique of this reformulation.

### iii. What types of preference are valid in a deliberative democracy?

When first generation deliberative and aggregative models of democracy are juxtaposed to each other, one of the clearest points of departure is the nature of the preferences that each articulation uses for collective decision-making. Aggregative models deal with preferences based on raw self-interest, deliberative models on preferences that are considered and reflect the public good. But if other forms of communication are allowed into the model, which explicitly draw upon personal history and experiences, it begs the question of whether self-regarding preferences and opinions should also be included. The challenge is therefore whether deliberative models of democracy should in fact relax the requirement that collective decisions should be made on arguments (and votes) that reflect the common good, and instead incorporate other types of preferences as legitimate inputs.

It is almost ubiquitous for second-generation deliberative democrats then, to relax the requirement and discuss preferences as ‘taking into account the perspective of others during the process of judgement’ (Niemeyer 2004; p. 352), or put more explicitly, in terms of a ‘public spirited attitude’ (Chambers 2003; p. 318).<sup>11</sup> But some go even further still. Mansbridge et al (2010) for example, argues for the principle of self-interest as both a necessary and even desirable feature of deliberative democracy. On the grounds of necessity, they follow Cohen and Rogers (2003) in arguing that even in an idealised first generation articulation, expressions of self-interest are required so that participants can gain a sense of the ‘common good’. Every individual involved in the process needs to know how a particular policy might affect every other individual in order to establish what is best for everyone. Without such an exploration, the challenge of Sanders (1997) becomes relevant, as ‘the understandings of the common good of the more powerful in the polity will dominate’ (Mansbridge et al 2010; p. 74). They also, more controversially, argue for the role that self-interested preferences play in the construction of an aggregate conception of the common good. This last point runs nicely into the second set of challenges that second-generation deliberative democrats have grappled with.

#### *Difference Challenges to Deliberative Democracy*

---

<sup>11</sup> The various ways this is expressed in second-generation models suggests a number of different interpretations as to exactly what these ‘public spirited’ preferences are. Although I do not want to get into the debate here, this forms one of the central discussion points in next chapter of the thesis.

The next challenge and set of adaptations that I want to consider draws on this idea that self-interest might have a legitimate place in deliberative democracy. I want to look at how concerns over the inevitable effect of deep disagreement have impacted on the movement from first to second-generation models. Both Habermas and Rawls point towards the power of reasoned discussion to achieve consensus – either via public reason, or through communicative rationality. But a number of different deliberative democrats (Bohman 1995,1996; Gutmann and Thompson 1996, 2004; Manin 1987) have identified examples where even under these principles, agreement is impossible.<sup>12</sup> Moreover some have even gone further, suggesting that complete consensus (where all individuals agree to the same outcome for the same reasons) is ‘unnecessary, and undesirable’ (Dryzek 2000; p. 170).

More specifically, they point to cases of value pluralism, where ‘reasonable’ preferences are completely contradictory to each other, and no common ground or ‘best for everyone’ decision is possible. Rawls' (1993; p.243) example of abortion, for instance, can be used for an effective illumination of this issue. Appealing to 'three important political values: the due respect for human life, the ordered reproduction of political society over time... and finally the equality of women as equal citizens', he claims that at the start of pregnancy the political value of female equality will have primacy, and therefore supports the right to abortion. But what if people think, also on reasonable grounds, that due respect for human life is more important? This position can also be defended on reasonable grounds. Herein lies the problem. Value pluralism dictates that two polar positions can both seem reasonable from different perspectives (Dryzek and Niemeyer 2006).

Once again, partitioning off the discussion of how this deep disagreement (a fact highlighted in second-generation accounts) might consequentially affect the nature and definition of deliberation, two resultant challenges need to be considered.

#### iv. What outcome does deliberative democracy yield?

Accounts of deliberative democracy that accept (and even embrace) value pluralism must therefore rely on outcomes other than a strict consensus. Mansbridge et al (2010) outline three different types of agreement that might be reached in a deliberative

---

<sup>12</sup> These approaches are often referred to under the heading of ‘difference democrats’.

democracy other than convergence (which they define as strict consensus). The first draws on the work of Sunstein (1996, 1997), who in studying decisions made by the US Supreme Court (often held up as an exemplar of deliberative decision-making), identified forms of workable agreements that continually came up. Termed ‘incompletely theorised agreements’, these outcomes reflect a group of individuals who all support the same substantive policy decision, but each for completely different reasons. Inserting this objective as a legitimate outcome for a deliberative democracy is easy, Dryzek (2000; p.48) argues, so long as the relative reasons that are used to sustain the agreement withstand appropriate scrutiny. The second possibility Mansbridge et al (2010) identify, they term ‘integrative negotiation’, and is defined as an agreement where individuals are able to exploit the incommensurate nature of the reasons they disagree over a particular decision. Whilst the third, cooperative negotiation, is based on the idea of compromise – where each individual ‘gives up’ some part of his or her preferred outcome in order to reach an agreement.

Other second-generation theorists have also explored some other types of possible outcome as the objective for deliberative democracy. Dryzek and Niemeyer (2006; p. 638), for example, identify three different levels at which ‘consensus’ might work:

	Value	Belief	Expressed Preference
Type of consensus	Normative consensus	Epistemic consensus	Preference consensus
Meta-counterpart	Recognition of types of value that are legitimate	Acceptance of credibility of disputed beliefs	Agreement on the nature of the disputed choices

Figure 1.2: Elements of Preference Construction

Very briefly, normative consensus concerns reaching an agreement over the ranking of values that are legitimately held by individuals. Epistemic consensus on the other hand, refers to an outcome where individuals agree on exactly how the disputed choices will impact upon the individuals. Preference consensus, of course, is strict agreement over the policy choice (which includes both strict consensus as well as an incompletely theorized agreement). They then extend this directly to the question of deliberative democracy, and argue particularly for a reformulation of the theory to focus on the ideal outcome as one of meta-consensus. Agreement over the set, rather than ranking, of values that are deemed legitimate, or the set of credible beliefs over the impact of the

policy alternatives are two possibilities, as is an agreement over the set of possible policy choices put before the demos. This third level in particular, has proved popular with social choice theorists who have explored the likelihood of preference structuration (Dryzek and List 2003; Miller 1992) - and is something I will explore in depth during chapter two when I consider the rational choice approach to deliberative democracy. Finally, Niemeyer and Dryzek (2007) also explore an idea they call 'intersubjective rationality', which argues for a strengthening of the link between values/beliefs and expressed preferences as the ideal outcome (in apparent contradiction to Sunstein's incompletely theorised agreements).

#### v. How are the outcomes in a deliberative democracy identified?

Deliberative models of democracy, even if they include other forms of participation as acceptable in their framework, are still clearly 'talk centric'. Amongst relatively small deliberating groups of individuals, it is perhaps possible to imagine that towards the end of the discussion, members will begin to get a sense of the likely outcome. For example, if each individual makes exactly the same reasoned argument towards the same policy proposal, it would be quite clear from the discussion alone that a complete consensus had arisen. Indeed if you consider small committee meetings, it is often the chairperson's responsibility to identify the point at which this happens. It is this sentiment, then, that seems to be the impetus behind statements like 'decision making by discussion' (Elster 1998c; p.1).

However, a number of political theorists and deliberative democrats alike have identified problems with this idea, and can be split broadly into one of three subcategories. Firstly then, Manin (1987) and Cohen (1997) are typical of first generation deliberative democrats in identifying consensus as an ideal, but both admit the possibility that when it is not forthcoming, 'deliberation concludes with voting, subject to some form of majority rule' (p. 75). This is very much an argument of second best. Voting is only required when consensus is not forthcoming. Other democratic theorists, particularly Saward (1998), have taken a different approach and argued for aggregation on the grounds of inevitability. Even under full agreement, they argue, some form of voting is required to fully clarify/identify the decision that has been made:



No matter how much deliberation takes place, heads have to be counted – aggregated – at some point if a democratic decision is to be reached. No adequate model of democracy can fail to be “aggregative”.

(Saward 1998; p.64)

The third subcategory involves a more positive approach to the challenge, and includes arguments that make the case directly for the benefits that aggregation – voting in particular – brings to a deliberative democracy. Prezeworski (1998) for example, claims that aggregation is the true site of legitimate authority, and that without it deliberative models of democracy are less complete. Goodin (2008b) on the other hand, takes a slightly different tack. He argues that deliberation is a highly path dependent process. Because of its serial nature, and accepting some level of epistemic virtue in the model, he asserts that aggregation’s simultaneous structure is a much better ‘decision-making’ process. In effect, it prevents the possibility that a ‘good’ choice is impacted by ‘bad’ choice made before it. Deliberation is essential for democracy, but it must end in voting.

My recommendation is therefore, ‘first talk, then vote’. That is to say, build discursive and deliberative elements centrally into the political process, but make the ultimate decisions through more purely aggregative procedures.

(Goodin 2008b; p.124)

Let me summarise some of the key points from these discussions then. Second-generation models of deliberative democracy have attempted to respond to a number of challenges to the original articulation in the work of Habermas, Rawls and Cohen. These challenges, roughly split into ‘exclusivity’ and ‘difference’ problems, have in turn led to theories of deliberative democracy that include of other forms of participation, communication and preferences, and have recast the objective outcome into one of a number of different possibilities, all identified by an aggregation phase. But what I have continually left to be discussed, of course, is how the concept of ‘deliberation’ has been reformulated to accommodate these challenges. This is the debate I now take up.

## **1.3 DEFINING DELIBERATION**

---

### **1.3.1 FIRST GENERATION DELIBERATION**

Deliberation and deliberative democracy are not the same thing. Nor are they bound together in perpetuity. It is possible to have democratic decision-making without any deliberative element, and similarly, it is quite plausible to have deliberation outside of a democracy. I make this point again because it will prove crucial for the concluding chapters of the thesis. As I have shown then, second-generation models of deliberative democracy depart from their first generation predecessors on a host of different dimensions. The final aspect I want to consider is without doubt the most important of these many departures. As Rawls (1997b; p.772) once stated, ‘the definitive idea for deliberative democracy is deliberation itself’. As a concept in its own right, deliberation is credited in delivering numerous benefits other than increased legitimacy in democratic outcomes. Mill, for example, famously argued for its epistemic and developmental qualities when he claimed:

No wise man ever acquired his wisdom in any mode but it [deliberation]; nor is it in the nature of human intellect to become wise in any other manner. The steady habit of correcting and completing his own opinion by collating it with those of others, so far from causing doubt and hesitation in carrying it into practice, is the only stable foundation for a just reliance on it [...].

(Mill 1972; p.88)

Other than accepting the key assumption that deliberation causes individuals to change through the process, I want set aside the discussion of exactly how, until the following chapter. Presently, my objective is simply to arrive at an adequate definition of what is meant when the concept of deliberation is utilised in deliberative models of democracy. There are then, two ways of describing this concept. One is by looking at the various tasks that deliberation is required to perform, which I will term the ‘structural account’. The other involves looking at the actions of the individuals involved in the process (or features of the communication) – which I will term the ‘behavioural account’. Political theory has remained fixed almost entirely on the latter, and it is this approach that I discuss first. To do this, I want to draw upon a three-fold framework outlined in Thompson (2008; p. 501):

Conceptual Criteria	Evaluative Criteria	Empirical Criteria
Clarifies the limit and scope of the concept.	Clarifies exactly what is meant by the concept.	Clarifies the concept in terms of operational features.
Defines what deliberation is applied to.	Defines the criteria by which to judge it.	Required for real-world applications

Figure 1.3: Three Levels of Deliberative Criteria

First generation approaches to deliberation offered a highly normative and idealised account of exactly what it meant. Recall once more Rawls' reliance on the 'give and take of public reason', or Habermas' appeal to 'fully rational discussion that requires individuals to deliberate with reference to an ideally inclusive community'. The objective of both these accounts, I would suggest, was to bridge the gap between conceptual and evaluative criteria. On the Rawlsian side, this is most clearly articulated in Cohen's (1997; pp.73-75) principles of ideal deliberation:

- Ideal deliberation is free in that individuals consider themselves bound only by the results of the deliberation, and are free to act on the decision.
- Ideal deliberation requires individuals to offer reasons for their proposals, on the requirement that the reason alone is sufficient to persuade others of its merit.
- Ideal deliberation requires that all individuals involved in the process are formally and substantively equal.
- Ideal deliberation aims to arrive at a rationally motivated consensus.

Whilst for Habermas, these might be expressed as the following (cited and adapted from Bächtiger et al 2010; p. 36):

- No one with competency to speak and act may be excluded from discourse.
- All have the same chances to question and/or introduce any assertion whatever as well as express their attitudes, desires and needs.
- No one may be prevented, by internal or external coercion, from exercising these rights.
- All have the right to question the assigned topics of conversation.
- All have the right to initiate reflexive arguments about the very rules of the discourse procedure and the way in which they are applied or carried out.

- The discourse must be public and rational, with participants adopting arguments that are orientated towards a truthful common understanding.

The critique these criteria have faced, as with first generation models of deliberative democracy, are challenges on a number of different fronts. As well as the problems that might be subsumed under the headings of exclusivity and difference challenges, there are also clear cognitive and motivational issues to consider. Do individuals really have the ability or even desire to participate in deliberation defined by these principles? Because of this, second-generation deliberative democrats have sought to redefine the notion of what is meant by deliberation. However, most have done so in a way that sits firmly in Thompson's (2008) first box. They have established conceptual criteria for a reformulated definition, but have then failed to translate this to an evaluative approach. In other words, most second-generation versions of deliberation offer quite vague and non-specified accounts of exactly what they mean. For example, Chambers (2003; p. 309) defines it as 'debate and discussion aimed at producing reasonable, well informed opinions in which participants are willing to revise their preferences in light of discussions, new information, and claims made by fellow participants'. Dryzek (2000; p.2) adopts a more liberal position in claiming 'the only condition for authentic deliberation is then the requirement that communication induce reflection upon preferences in a non-coercive fashion'. Whilst Gutmann and Thompson (2004) fail to provide an overt single definition of what is meant by 'deliberation' – focussing purely on debates in deliberative democracy.

This is clearly a significant problem. If the effects of deliberation are to be studied in the real world, in order to provide insight and analysis as to how it is then integrated with normative democratic theory, then a clear definition is vital. How can something be investigated if it cannot be identified in the first place? This is a sentiment echoed by Steiner's (2008) warning regarding the apparent prevalence of 'concept stretching' in the discipline. If deliberation includes everything, then by definition, it refers to nothing. It is then impossible to tell whether an empirical study is examining the effects of deliberation, or another related form of communication such as cooperative argumentation (Makau and Marty 2001), dialogue (Pearce and Littlejohn 1997) or debate (Tannen 1998).

### 1.3.2 DEFINING 'SECOND-GENERATION' DELIBERATION

In this section I want to construct a model of deliberation that draws on a mixture of the principles identified by first generation theories, but makes explicit the possibility of relaxation on the basis of the various challenges made to them. In many respects then, this is a similar approach to that taken by proponents of the Discourse Quality Index (Sporndli 2003, Steenbergen et al 2003)<sup>13</sup>, and indeed I will draw quite heavily on their categories in my discussion of the principles of deliberative behaviour. But where my approach differs, is that I also want to provide a discussion of deliberation in terms of *how it proceeds*. By looking at both ways to define deliberation, my objective is to create a more comprehensive set of evaluative criteria that can be used to both judge its quality, and particularly, to justify empirical/operational features of the deliberative mini public that I outline in chapter five.

But before I do this, I need to make a fundamental point very clear. It is possible to define concepts in absolute terms. For example in chemistry, an element is either hydrogen, or it is not. It cannot be ‘hydrogen-ish’. Following Sporndli (2003) and Steenbergen et al (2003), I argue that deliberation is not one of these notions. It exists on a scale. At one end sits something that satisfies the principles of deliberation perfectly, and at the other end, something that fulfils none of these criteria. Of course, this does not mean that points on the scale of deliberation (or its constitutive principles) cannot be identified and labelled as a particular form. When we deal with percentages in mathematics, for example, we continually rely on terms that pertain to specific points on this scale – 50% is often described as ‘half way’. Neblo (2007) makes a useful related point when he discusses the same premise, but rather concerns himself with the point at which discussion scores so low on the scale that it does not warrant the label deliberation at all:<sup>14</sup>

---

<sup>13</sup> The ‘Discourse Quality Index’ (DQI) (Sporndli 2003, Steenbergen et al 2003; pp. 27-30) draws on five ‘coding categories’ in order to assess the extent to which real world deliberation satisfies the Habermasian inspired ideal. The first is participation, which includes the extent to which individuals can participate freely in the deliberative process. The second is the level of justification, which concerns the extent to which participants offer reasons for their positions. Related to this is the third criterion, which looks at the way these reasons are justified, from self-interest to the common good. The fourth is respect, which measures the degree to which individuals treat each other in terms of counterarguments, and the final aspect is constructivism, which analyses whether the deliberation is directed towards the pursuit of an outcome amenable to all involved.

<sup>14</sup> Neblo (2007) begins with a very useful articulation of these issues, on the promise that in the latter section of the paper he will explore the principles/criteria he believes are useful in order to define and judge deliberation. I find his exposition extremely unstructured, and because of this, I’m not really sure he achieves his objective.

Deliberation is thus akin to a concept like ‘courage’, that describes a range of phenomena, but does so in a way that is intrinsically approbative. There may be degrees of courage, but we need a different phrase, ‘utter cowardice’, to describe a complete lack of courage. Similarly, putative deliberation that falls below a certain threshold is no longer deliberation.

(Neblo 2007; p. 529)

I now want to outline the six principles that I believe constitute the main normative features of deliberation from the behavioural approach. In doing so, I also want to explore the way some of these criteria have been relaxed, and how this relates to the ‘quality’ of deliberation. These six criteria are:

- i. Interactive Communicative Process
- ii. Equality
- iii. Mutual Respect and Reciprocity
- iv. Reason-based Discussion
- v. The Public Principle
- vi. Decision-focussed

Figure 1.4: The Behavioural Account of Deliberation

#### *Interactive Communicative Process*

Deliberation, ideally, is an interactive communicative process between two or more individuals (Minozzi et al 2010). This point is made quite explicit by Habermas, when he claims ‘moral justifications are dependent on argumentation actually being carried out, not for pragmatic reasons of an equalisation of power, but for internal reasons, namely that real argument makes moral insights possible’ (Habermas 1990; p. 57). It might seem to sit in apparent contradiction to some interpretations of the Rawlsian approach, which places more emphasis on internal reflection. Those who subscribe fully to this account, thus, might define deliberation as something that ‘occurs anytime a citizen either actively justifies her views (even to herself) or defends them against a challenge (even from herself)’ (Gunderson 1995; p. 199). In other words, it is possible to take part in deliberation purely with oneself. There are others, however, who whilst recognising the positive deliberative impact that such a cognitive process yields, recognise that it cannot entirely replace the ‘external-collective’ dimension (Goodin 2000). Challenges especially relevant to deliberation articulated in this way, including

the ability to make others present without having met them, or the ability to then understand their arguments without having heard them, all clearly suggest that some element of actual communicative interaction is required.<sup>15</sup> As Shapiro (2002; p. 197) pointed out: ‘we can be individually reflective, but not individually deliberative’.

A second way that the maxim of interactive communication might be challenged and relaxed surrounds the physical nature of the contact. In particular, the possibility that information technology might ease the burden of face-to-face deliberation has been explored by a number of second-generation theorists (Coleman and Gotze 2001, Dahlberg 2001, Smith et al 2009). Wright and Street (2007), in a discussion of the institutional factors that play a role in determining quality of deliberation online, identify a particular benefit other than addressing the scale problem that it might provide. They distinguish between synchronous and asynchronous forms of communication, where in the latter participants have an opportunity to go away and think about their response before making it, compared to the former where it must be instantaneous. Clearly, individuals who take more time in considering their response are behaving more deliberatively. But this isn’t precluded, per se, in synchronous behaviour. Wright and Street’s (2007) argument is indeed interesting and important, but their assertion that it does not occur in face-to-face deliberation is erroneous. Gastil (2000), on the other hand, considers this question directly, and asks what virtues face-to-face interaction delivers that is not provided via web-based activity. Although he openly admits the case isn’t quite settled, he cites the importance of socialisation and group cohesiveness in making decisions over controversial political decisions - something facilitated by face-to-face discussion.<sup>16</sup> In short then, whilst deliberation is indeed possible online, and even within a single individual, face-to-face interactive communication remains the ideal.

### *Equality*

The principle of equality is a feature common to almost every comprehensive definition of deliberation, and appears particularly when discussed with reference to deliberative democracy. Cohen (1998), for example, identified two levels at which it is relevant, the procedural and substantive:

---

<sup>15</sup> Goodin (2003; p. 108-109) also identifies the issue of legitimacy as a reason to defend the need for physical communication between individuals.

<sup>16</sup> This is an area that I will discuss in much more depth during chapter three.

They are formally equal in that rules regulating the ideal procedure do not single out individuals for special advantage or disadvantage. Instead, everyone with deliberative capacities – which is to say more or less all human beings – has and is recognised as having equal standing at each stage of the deliberative process. [...] The participants are substantively equal in that the existing distribution of power and resources does not shape their chances to contribute to deliberation, nor does that distribution play an authoritative role in their deliberation.

(Cohen 1998; p. 194)

First generation accounts of deliberation have therefore defined it as discussion that takes place between individuals who are both substantively and procedurally equal. However, second generation accounts have had to respond to significant criticisms on the first level. Sanders' (1997) argument regarding power structures has raised the possibility that a completely equal deliberation is impossible, and that individuals with greater deliberative capacity will dominate the discussion. Moreover, the desirability of such a concept in particular debates is also questionable. For example during deliberation on complex ethical questions in the scientific realm (i.e. genetically modified foods), a persuasive argument that 'experts' might indeed warrant unequal status can be made. The relaxation of the equality maxim in deliberation has therefore mirrored Sen's (1992) famous reformulation of the economic approach to poverty and inequality, moving to a capability-based approach. In this sense, deliberative equality has come to rest simply on a procedural basis (Christiano 2008). Deliberation represents a process that values higher levels of equality in participation (the more equal the better the deliberation), but can only provide the institutional rules to guarantee equality of opportunity.<sup>17</sup>

### *Mutual Respect and Reciprocity*

Linked to the concept of equality are the maxims of mutual respect and reciprocity (Gutmann and Thompson 1996). Deliberation is a dynamic process, which requires individuals to interact with one another during the discussion. This means, of course, that they are therefore continually put in the position where they must respond to statements and claims made by other members of the deliberation. Mutual

---

<sup>17</sup> A point that has yet to be even acknowledged by deliberative democrats working on the topic of equality concerns the fact that 'whoever speaks first' automatically has disproportionate power compared to the other individuals involved. This is because they have the opportunity to set the agenda of the discussion. Because of the dynamic nature of deliberation, this contradiction seems endemic, although my discussion of the external account of deliberation might suggest one way in which it might be mitigated.



respect/reciprocity, then, are defined by Gutmann and Thompson (2004) in two parts. The first concerns ‘a favourable attitude towards, and constructive interaction with, the persons with whom one disagrees’ (2004; p. 79). This links nicely with the maxim of equality, and indeed Sanders (1997) continually merges these points during her critique. Individuals deliberating with each other must respect one another to the extent that they respond directly to the arguments and claims that are made during the discussion. Talking at cross-purposes, where a given participant pays no attention to what others are saying, occurs when this is violated and is clearly non-deliberative in nature.

The second part of the definition of this maxim is slightly more complex. Reciprocity concerns not just the requirement that individuals respond to the claims made by others, but also that they give them equal standing to their own. By this, it refers to what Habermas first alluded to when he used the now famous phrase ‘the forceless force of the better argument’ (Habermas 1999; p. 332). Gutmann and Thompson have then placed further emphasis on it, and described it as ‘the character of individuals who are morally committed, self-reflective about their commitments, discerning of the difference between respectable and merely tolerable differences of opinion, and open to the possibility of changing their minds or modifying their positions at some time in the future’ (2004; p. 79). Reciprocity, then, is required for deliberation because it explicitly includes the assumption that the process has an effect on the individuals involved. Without this foundation, there is no reason for deliberation in the first place (Mansbridge et al 2010; p.78). Better quality deliberation therefore involves individuals who are better able to meet this requirement.

#### *Reason-based Discussion*

The requirement that deliberation proceeds on the basis of reason-based discussion is a feature of all first generation accounts. Habermas focused on the contestation between rational discourses, whilst Rawls continuously referred to ‘public reason’. Reason-based discussion, understood here as linguistic/communicative device that simply draws a link between justification and an action, has a number of normatively attractive features that make its prioritisation in deliberation quite logical. Most importantly, the claims made on the basis of reason-based argument are much more likely to stand up to deliberative contestation, as well as persuade others of the merits of the argument. They are also, by definition, more universal in the sense that individuals with a requisite cognitive ability, irrespective of personal experience, are more likely to be able to

understand them. Finally, reason-based discussion also has the benefit that it invites commensurable replies from other participants, in that a proposition put forward by any individual can be countered by any other without having to establish a shared personal perspective/history<sup>18</sup>. As Manin (1987) put it then:

Between the rational object of universal agreement and the arbitrary lies the domain of the reasonable and the justifiable, that is, the domain of propositions that are likely to convince, by means of arguments, whose inclusion is not contestable, the greater part of an audience.

(Manin 1987; p. 363)

Because of these features, first generation models of deliberation took great pains to exclude other forms of communication in their definition. But many second-generation accounts have taken a more liberal position.<sup>19</sup> Fearon (1998), for instance, argues for the broadest conceptualisation of what should be included, and suggests deliberation should be defined merely as ‘discussion’. Dryzek (2000), on the other hand, takes a position somewhat in between these two perspectives, and claims there are some particular forms of discussion that should, and some that should not, form part of a comprehensive definition of deliberation. In Dryzek’s own words, he states ‘we must determine the degree to which [it] must stress rational argument, and the extent to which it can and should admit other forms of communication’ (2000; p. 67).<sup>20</sup>

The premise then, is the cognitive challenge made by Sanders (1997; p. 348) that ‘some citizens are better than others at articulating their arguments in rational, reasonable terms’, and that this power discrepancy means the deliberative process is likely to favour a particular type of individual. Sanders draws heavily on the work of Schumpeter, and in particular, his assertion that the masses ‘are not capable of rational argument’ (p. 354), and whilst her line of reasoning explicitly concerns the dominance of white middle class men during American jury deliberation (pp. 362-369), the point is certainly generalisable. Two possible avenues exist in order to confront this challenge

---

<sup>18</sup> Bohman and Richardson (2009; p.254) contest this slightly, and suggest that individuals have different conceptions of what counts as a reason. However, their argument appears to be made on the basis of the content of these reasons, rather than the linguistic/communicative structure of what a reason contains. This is addressed in the following criterion.

<sup>19</sup> A good discussion of how a related discipline, communication studies, has responded to these challenges is found in Escobar (2009) – on the topic of ‘dialogue’ rather than ‘deliberation’.

<sup>20</sup> Because Dryzek is typical of most second-generation accounts and fails to continually distinguish between the two concepts, he is actually referring to ‘deliberative democracy’ in this quote. The point is, though, must more applicable to the concept of ‘deliberation’.

to defining deliberation. The first is for deliberation to be recast so that it includes, or acknowledges a ‘training phase’ in the process, where individuals are able to learn to take part in talk that is ‘rational, constrained, and oriented to a shared problem’ (p. 370). This would ensure equality whilst protecting the privileged position of reasoned argument as the only form of communication valid in a deliberation.<sup>21</sup>

The second possibility, one Dryzek himself embraces, is to actively include other forms of communication into the model:

Some deliberative democrats, especially those who traffic in ‘public reason’, want to impose narrow limits on what constitutes authentic deliberation, restricting it to arguments in particular kinds of terms. A more tolerant position, which I favour, would allow argument, rhetoric, humour, emotion, testimony or storytelling, and gossip.

(Dryzek 2000; p. 1)

His criterion for inclusion, then, is two fold. On one level, it is based on a distinction between types of talk that cause individuals to reflect on their preferences in either a coercive, and non-coercive manner. The latter, including manipulation, indoctrination, propaganda, deception, and threats, offer no benefits to individuals involved in the process and should therefore be excluded. On the other hand, rhetoric (Dryzek 2010), emotion (van Stokkom 2003), testimony (Mansbridge 1990, Sanders 1997) or storytelling (Black 2008; Ryfe 2006) each has various positive reasons to be included, beyond the fact that they represent much easier ways for ‘everyday’ people to participate. The second level is based on a requirement that any form of talk must bridge the gap between the specific and the general. Just as reasoned argument is able to do this, so should any other form of communication if it is to be included in the definition of deliberation. Rhetoric, for example (Dryzek 2010), is claimed to perform extremely well on this dimension, as it has the potential to help individuals understand the issues and arguments presented by minorities. Emotion, testimony and storytelling, whilst highly personal in nature, also often appeal to universal principles. Fables, for instance, are extremely context specific yet convey a message that is applicable to almost everyone in society. These points, Dryzek suggests, are enough to support their inclusion.

---

<sup>21</sup> Although Sanders (1997) might argue this equates to indoctrination into a pre-determined hierarchy.

So where does this leave a definition of deliberation? To include these other forms of communication into the formulation raises the question of prioritisation. Whilst individuals might indeed utilise these different types of talk at different moments (something I will discuss in the structural account below), the more reasoned discussion that takes place, the better the quality of the deliberation. If no reasoned discussion takes place whatsoever, it reaches the point that Neblo (2007) identified: it is not deliberation.

### *The Public Principle*

The public principle, then, is related to the previous criterion. If being ‘reason-based’ is a procedural maxim that refers to how speech should be constructed in a deliberation, then the public principle concerns the substantive aspect: what the reasons should include or be based upon. First generation accounts of deliberation, particularly Rawls, relied on the principle of ‘public’ reason defined as the following:

[...] its subject is the good of the public and matters of fundamental justice; and its nature and content is public, being given by the ideals and principles expressed by society’s conception of political justice, and conducted upon to view on that basis.

(Rawls 1997a; p. 93)

The argument was simple. Deliberation, if the objective was to arrive at a consensus, should rely on reasons that are substantively applicable to all the individuals participating. This is clearly best achieved by referring to a conception of the common good that is supported by all individuals. Further to this, it is also best ensured by making sure that discussion takes place in public, and draws upon language that makes this collective nature explicit. But as I have discussed above, the idea of a single conception of the common good cannot be guaranteed, a fact that difference democrats have used to challenge the objective of consensus. In these cases, it impossible for individuals to appeal to a substantive point to generate ‘reasons that all can accept’ (Bohman and Richardson 2009). The slightly relaxed position, one most second-generation accounts adopt, is to then rely on reasons that ‘the public at large could accept’ (Chambers 2004; p. 390). This, it is argued by Mansbridge et al (2010), then

also allows for the possibility that statements on the basis of self-interest might be accepted as both valid and beneficial to deliberation.

Furthermore the public principle also includes Rawls' second aspect in his definition, often referred to as the Kantian derived 'publicity principle'. The fact that it operates in public, and reasons are put forward in language that takes this into account, has been relied upon by a number of second-generation perspectives to deliver a positive outcome in the face of pluralism. Gutmann and Thompson (2004; p. 135), for instance, claim that 'the principle of publicity requires that reason-giving be public in order that it be mutually justifiable', whilst Elster (1997; p. 12) makes a further claim that the requirement of making arguments in public will eventually lead to one becoming swayed by these considerations.<sup>22</sup> To sum up then, whilst deliberation clearly requires some element of self-interest at particular times in the process it should indeed prioritise arguments that are based on a public principle.

#### *Decision-focussed*

As a concept in its own right, deliberation has been used in a number of different contexts (Fung 2007). Some of these, for example Mill's articulation in terms of education and personal development, require only the process to achieve their desired ends. Nothing need happen other than individuals participating in the deliberation. But at this point, I want to slightly backtrack to the first half of this chapter. Deliberative models of democracy are designed to show how collective decisions can, and should, be made by a society. Because of this, the maxim of being decision-focussed is explicitly relevant to formulations of deliberation when used in this context. Indeed, this is why those working on deliberation in areas other than democratic theory continually rely on the term 'democratic deliberation' (Barge 2002, Gastil 2000). The thrust behind the claim that deliberation must include 'decision-focussed' discussion is therefore more one of necessity: without some desire to come to a conclusion, it is possible for deliberation to have no real end point. As a maxim for deliberation, then, the maxim of being decision-focussed is unique in that it only really refers to the latter stages of the process. In this sense, whilst I include it in the list of behavioural criteria, it occupies a unique position in that it is only really relevant when taken in conjunction with the structural account of deliberation. It is this approach that I now turn to.

---

<sup>22</sup> Chambers (2004) and (2005) provides a nice critique of these assumptions, and argues that in some cases, the publicity principle might be harmful for deliberation.

### 1.3.3 THE STRUCTURAL ACCOUNT OF DELIBERATION

The behavioural approach to deliberation, on its own, offers a description of the nature of the discussion that takes place during deliberation. But it says nothing as to what exactly is being discussed, and at what point. It is, if you like, the difference between the question “what *is* deliberation”, and the question “*how* does a deliberation take place”. This alternative way of looking at the concept is fundamental to devising an empirical test of the process at work. Not only does it provide a loose structure that can be drawn upon to establish key subtasks that should be completed in the overall deliberation, but it also offers a logical suggestion for the order in which these should be done. Some first generation deliberative theorists in political science have indeed pointed towards this debate as important, for example Cohen (1997) writes:

There are three general aspects of deliberation. There is a need to decide on the agenda, to propose alternative solutions to the problems on the agenda, supporting those solutions with reasons, and to conclude by settling on an alternative.

Cohen (1997; pp.73)

Whilst a second-generation definition of deliberation in terms of political theory has been adapted to include, for example, other forms of communication and other justifications for reasons in the model, on its own it fails to then discuss the times at which these respective relaxations of behavioural principles are appropriate. The same cannot be said for the field of communication studies, where this approach has received significant attention (Gastil 1993, McLeod et al 1999, Pearce and Littlejohn 1997). I now want to outline one such articulation, before discussing briefly the normative arguments for defining deliberation in this way. I start, then, with a definition of deliberation that is rooted in what Gouran and Hirokawa (1996) term the ‘functional theory of group-decision making’:

[...] full deliberation includes a careful examination of a problem or issue, the identification of possible solutions, the establishment or reaffirmation of evaluative criteria, and the use of these criteria in identifying an optimal solution.

(Gastil 2000; p. 22)

This has been further expanded in Burkhalter et al (2002), who offer an account of deliberation that is rooted on four key phases. Each represents a specific ‘task’ that must be completed by a deliberating group within the process:

- i. Education and Information Phase
- ii. Identification of Solutions Phase
- iii. Evaluative Criteria Phase
- iv. Decision-making Phase

Figure 1.5: The Structural Account of Deliberation

#### *The Education and Information Phase*

The information and education phase is defined as the point in a deliberation that is specifically targeted towards ensuring individuals understand the problem they are presented with. In this sense, it involves the presentation, clarification and discussion of evidence. Higher quality deliberation relies on the accuracy and comprehension of such knowledge. Relating this to the behavioural approach for one moment, it is clear that relaxations of the maxims of reason-based discussion and the publicity principle are perhaps more valid at this point than at any other in the deliberation. The sharing of personal experiences through devices such as story telling or personal testimony are without doubt effective devices of discovery, especially when the individuals involved are those impacted upon by the issue.

#### *The Identification of Solutions Phase*

The second element of a deliberation, once all participants have gained a thorough understanding of the issue, concerns the identification of potential solutions to the problem. In some deliberations, a set of possibilities might be presented to the group before hand, whilst in others it might be completely open ended. In this second case, the deliberating individuals must work together to come up with a list of alternative policies on the basis of the information they received during the first phase. At a minimum, two solutions must be identified, although as Mill would argue, higher quality deliberation is reflective of a wider spectrum of possible outcomes.

#### *The Evaluative Criteria Phase*

Once a full set of potential solutions is identified, the next stage in a deliberation is for the participants to then set the terms of evaluation. One is the possibility that

individuals come to agree on a single perspective by which to rank the outcomes, which also links with the behavioural criterion of the publicity principle – i.e. ranking options of the basis of the common good, or at least according to one conception of it. The second possibility is that they arrive at a restricted domain of legitimate evaluative criteria, for example deliberative forums are often claimed to prioritise empirical or scientific approaches to a problem. The final possibility is the rather more minimal condition that individuals at least make their evaluative criteria public to the other members.

### *The Decision-making Phase*

Finally, the decision-making phase concerns the outcome of the deliberation, and requires that the group make a binding choice or recommendation. Within a deliberation, this might take one of two forms. Either it allows the individuals involved to make a group decision by way of discussion: for example in small groups it might simply be obvious to all involved that a position has been arrived at. Or in larger deliberating groups, this is more likely to take the form of an aggregation mechanism – a show of hands, for instance. Whilst these two differ in their *modus operandi*, what they share is the fact that discussion in this phase is required to take on a much more decision-focussed element.

All four of these phases, I would argue, are crucial for any effective deliberation to take place. Without an information or education phase, for example, the individuals might make a decision that is based on incorrect assumptions or a misunderstanding of the issue. Or without an identification of solutions phase that emphasises pluralism, it is quite possible for the group to arrive at an inferior decision. The question that remains, then, is if these phases, (articulated as distinct points in a deliberation), should be undertaken by individuals in this specific order. Burkhalter et al (2002) favours a definition that does not require linear progression through them, although they point out that ‘significant revelations in more primary phases have implications for other phases (e.g. changing the evaluative criteria creates the need to revisit the evaluation phase)’ (p.420). Similarly, all second-generation accounts of deliberation conceptualise the process as a unitary forum, where overlap between these various phases, which are not seen as distinct, is possible. The very fact that individuals are able to revisit earlier phases is an important element of what makes deliberation so different and so beneficial.



Whilst this is without doubt true, I would suggest an ever so slightly stricter interpretation, and identify a crucial normative reason as to why higher quality deliberation should at least *attempt* to follow this order. Or at the very least, that the education and information phase is given some form of independent recognition at the start of the process. A major pathology identified in the structure of deliberation has been the challenge that minority positions (and persons) present. When this is taken in conjunction with the argument of path dependency put forward by Goodin (2008), a serious potential pitfall is made much more likely. By definition majority perspectives (even if they are held by individuals adhering to the maxims of equality and mutual respect/reciprocity), have a much greater impact on the eventual decision than minority ones. This means that deliberation is susceptible to what Sunstein (2003) identified as the potential of polarisation, or in language that pertains specifically to the dynamic nature of deliberation, to a ‘polarisation cascade’ (Hamlett and Cobb 2006). Ensuring a thorough understanding of the issue at hand, and then formally investigating the possible solutions in a distinct phase *before* focussing the deliberation on the decision should mitigate this risk. In short, the structured approach aims to safeguard a key salutary benefit – that deliberation arrives at the ‘best’ outcome possible, irrespective of the nature of the individual(s) supporting it.

#### **1.4 CONCLUDING REMARKS**

---

To reiterate then, the first objective of this chapter has been to provide an outline of the way deliberation and deliberative democracy have responded and been recast on the basis of contestation and evidence: the movement from first to second-generation accounts. In doing so, the final section of the chapter included an outline of what deliberation entails, and provided a number of evaluative criteria that might be used to both judge, as well as promote, deliberation in an empirical setting.

From this foundation, in the next chapter of this thesis I want to examine a particular claim made by deliberative theorists working in all areas: that deliberation causes ‘revision’ and transformation of the individuals involved in it. As Mansbridge et al (2010; p. 78) state, ‘deliberation would have no point if it did not produce change in the views of at least some participants’. It is a consideration of this key question that I now address.

## — CHAPTER 2 —

# A RATIONAL CHOICE APPROACH TO DELIBERATIVE REVISION: PREFERENCE, ISSUE AND AGENCY

### 2.1 INTRODUCTION

---

In the previous chapter, I outlined the way in which the twin concepts of deliberative democracy and deliberation had responded to various challenges, detailing the shift from a first to a second-generation understanding. Whilst it is most often the case that deliberation is argued to impact positively on the basis of legitimacy, other theorists have sought other dimensions of justification. Arendt (1970), for example, particularly focuses on the educative power that political deliberation enjoys. Estlund (1997) on the other hand, proceeds explicitly on the grounds that deliberation acts as a sensitive truth-tracking device, delivering more ‘correct’ collective decisions.

In this chapter, I do not want to settle any of these debates. Deliberation may indeed produce better outcomes on all these different dimensions. Rather, my objective is to in fact add a layer of complexity to a claim central to all of the first and second-generation accounts of deliberation. Two key features of these models are firstly, that partial authority is handed over to the force of the better argument, and moreover, secondly that this authority is now understood as a dynamic concept. Something that requires justification within a process, however, is therefore also quite clearly susceptible to being challenged and, by extension, allows for the possibility that a revision might take place. At the very centre of the notion of deliberation, then, is the explicit and quite logical proposition that potential collective decisions are subject to the forces of change over the course of a sustained deliberation:

The core of the theory, then, is that rather than aggregating or filtering preferences, the political system should be set up with a view to changing them by public debate and confrontation.

(Elster 1997; p.11)

This sentiment expressed in Elster's now famous quote is often taken as a definitional starting point for deliberative theorists. Collective decisions are about the aggregation or construction of preferences, and therefore this is clearly the most important level at which any revision might take place. Whilst I do not dispute the claim that preference revision *is* important and requires investigation, it is though, not the only way in which deliberation can alter the nature of decision-making. Preferences are one element of a three-stage concept. They cannot exist without an issue upon which opinions can be formed. And when these are both present, the individual in question must then decide how to act upon them: a step between preferences and action that is often referred to as a 'mode of reasoning'. A combination of these three interrelated concerns, which I term issue, preference, and agency, thus provides the framework upon which the notion of deliberative revision should be built, applied and most relevantly, investigated.

This second chapter therefore has three key concerns:

- i. To construct an appropriate analytical framework in which to categorise and study the effects of deliberation on the individuals partaking in it.
- ii. To draw on the rational choice approach as a tool to unpack, and explicitly articulate, what is meant by the various deliberative claims of revision.
- iii. To show that whilst the notions of preference and issue revision are important, (and have been the focus of most empirical work in deliberative democracy), that a third concern, agency, should not be ignored.

Indeed this final point then provides the platform for the subsequent conceptual and experimental sections of this project. Once I have established agency revision as both a distinct and important research question, I will then consider how it might be investigated empirically in order to then re apply the results back to deliberative models of democracy.

## **2.2 DELIBERATION AND PREFERENCE REVISION**

---

### **2.2.1 RATIONAL CHOICE AND PREFERENCE REVISION**

Let me start with the most obvious level, that of preferences. The application of rational choice theory to the question of democracy is nothing new. Indeed ever since Downs' (1957) seminal work, political science has recognised the advantages of formal

modeling. Being able to restrict and manipulate variables in artificial conditions allows different hypotheses to be conceptually tested before an experimental approach is applied. It is imperative, however, not to conflate two related traditions in this single vein of study. One is that of social choice theory, which is concerned predominantly with studying how rationality can be applied to different mechanisms of aggregation. In this sense, it works almost exclusively at the aggregate level, and concerns itself only with the nature of the overall outcome. Early applications of this approach, then, included Black (1948) and Arrow (1951), who generated a host of different theoretical results that were utilised by Riker (1982) to famously critique wholly aggregate models of democracy. Building on this approach, Miller (1992) and Dryzek and List (2003) have then transposed the discussion to deliberative democracy, and have attempted to show how social choice theory might be reconciled with this alternative method of collective decision-making. I will return to these various arguments a little later in this section.

The second subset of rational choice theory I want to refer to is decision theory, which proceeds at the level of the individual and is therefore concerned much more with decision-making and human action. On this basis, it is the structure of preferences that becomes conceptually relevant in conjunction with the revision argument allied to deliberation. However, whilst social choice informed discussions have generated significant attention both conceptually and empirically, the same has not proved true for that of decision-theory. Indeed, Austen-Smith and Feddersen (2006), and List and Dietrich (forthcoming) aside, which both concentrate on the mechanism of revision itself rather than the end result, there is very little consideration of how these two approaches might be combined. This, of course, leaves the obvious question: why concentrate on the latter aspect of rational choice theory and not the former? Critics of the approach fervently believe that its reach should be restricted to the domain of the consumer, and that any foray into subject matters beyond this represents the imposition of a wholly inappropriate paradigm (Green and Shapiro 1994). This claim usually rests upon two fallacious assumptions that I quickly want to dismiss. The first is that decision theory is only applicable to instances of a fixed issue at stake, with fixed preferences over the different possible outcomes. Under such rigid convictions, the central deliberative claim of revision is therefore quite contradictory. Or to put it another way, a decision theory informed analysis of deliberation and deliberative democracy is impossible. As List and Dietrich (forthcoming) summarise:

[In classical models] a rational agent has fixed preferences over fundamental alternatives or outcomes, such as fully described states of the world, and any observed changes in his or her preferences over less fundamental alternatives, such as policy options, are purely information-driven: they are due to the fact that the agent has learnt new information about which fundamental outcomes are likely to result from these options.

(List and Dietrich *forthcoming*; p. 2)

But there is nothing within the foundations of decision theory, understood as the maximisation of utility, which results inescapably in this commitment. It is just that classical models are silent on the issue; with this silence interpreted by critics as saying it cannot happen. In the real world preferences can, in some cases should, and indeed do change up to and after a moment of decision-making.<sup>23</sup> What is important is that at the point where a decision is required, that preferences are fixed for that moment. Whilst only recently has this argument begun to establish itself in mainstream research under the umbrella term of evolutionary economics, the notion of adaptive preferences goes back as far as Elster's 'sour grapes' (1983; pp. 109-140). In it he discussed a number of different ways in which the process might work, for example:

- Adaptive Preferences: unconscious, and reversible.
- Preference Learning: unconscious and irreversible.
- Pre-commitment: by committing to prevent certain preferences in the future.
- Manipulation: other individuals shape your preferences for their own ends.
- Character Planning: conscious and irreversible, i.e. choosing training.
- Wishful Thinking: changes perception of the situation, not preferences.

Whilst a discussion of these types of revision in relation to deliberative theory is certainly worthwhile<sup>24</sup>, and indeed links directly to the formal treatment offered by List and Dietrich (*forthcoming*), it is not the focus of my argument. At this point, I am merely concerned with the fact that decision theory does not itself prohibit a central claim made in deliberative theory. Therefore, when deliberative theorists talk of the revision of preferences, from a decision-theory perspective at least, what they are

---

<sup>23</sup> Sugden (2006) provides a good discussion of the relevance of this to economic theory.

<sup>24</sup> For example a consideration of which type of preference change compares most accurately with deliberative democratic theory.

actually referring to is the process of ‘evolutionary preference formation’, or as Goodin (1986) calls it ‘preference laundering’.

Nonetheless, it is hard to deny that there are instances of preference change which the standard model has difficulties explaining. Sometimes agents do undergo transformations that go beyond information learning in the ordinary sense [for example] a capitalist businessman who, after surviving a plane crash, decides to devote his life to charity [...]

(List and Dietrich *forthcoming*; p. 2)

The second erroneous claim to the incompatibility of decision and deliberative theories concerns the issue of human motivation. Quite often, when political theorists are asked to consider rational choice approaches, it is the work of Smith (1976b) that immediately comes to the fore. In particular, his now famous discussion of the virtue of selfish behaviour relating to the hypothetical situation involving the butcher, baker and brewer (1976b; pp. 26-7). Smith argued that a group of entirely self-interested individuals, interacting with each other on the basis of pure self-interest could, due to comparative advantage, ensure maximum economic well-being. But where this (largely historically motivated) oversimplification runs into trouble, is that it completely disregards the alternative view expressed within Smith’s ‘theory of moral sentiments’ (Smith 1976a). In this other treatise, he also explicitly acknowledged other motivational factors within human nature; in particular the effect that an outcome has on other individuals (Sugden 2002).

How selfish ‘soever man may be supposed, there are evidently some principles in his nature which interest him in the fortune of others and render their happiness necessary to him though he derives nothing from it except the pleasure of seeing it.

(Smith 1759; p.1)

Further to this, the real world as well as experimental economics, provides an abundance of counterfactual evidence of non-selfish behaviour. Charitable donations made by city-dwellers to organisations such as the Royal National Lifeboat Institution (Sugden 1993) are extremely difficult to explain by appealing to self-interest; as are the contributions of money in public goods games made by agents in laboratory conditions (Andreoni 1990, 1993). But again, decision theory, understood as the maximisation of

utility, is not irrevocably tied to self-interest: it simply involves the assumption that individuals choose the outcome that satisfies their preferences, whatever these are based upon. It is this particular issue, the nature of post-deliberative preferences, which I want to now consider.

### 2.2.2 PREFERENCES AND DELIBERATIVE REVISION

Deliberative theory's fundamental assumption that a political deliberation should change preferences is clearly ambiguous without any sense of the direction this movement should take. As I discussed in chapter one, this question relates directly to an issue at the core of the movement from first to second generation accounts. First-generation deliberative democrats, then, argued strongly for a reliance on consensus. Rawls (1993, 1997a, 1997b) and Cohen (1997) appealed directly to the notion of public reason, whereby all individuals would eventually possess preferences reflective of the common good. Habermas (1987, 1990, 1994, 1996a, 1996b, 1999), on the other hand, favoured an articulation in terms of a 'rationally motivated consensus'. But as I then demonstrated via the example of abortion and the 'incompleteness of public reason' (Schwartzman 2004), a reliance on the notion of the 'common good', and the associated consensus that normative political theorists have claimed then follows, has proven a rather idealistic and at times quite excessive claim for many deliberative democrats. Value pluralism, as Dryzek (2000; p.170) argued, was not only inescapable but beneficial for democratic decision-making.

A rejection of consensus around a single preference position as a goal for deliberation does not mean, however, that a mechanism of preference revision cannot take place. And more importantly, it does not mean that a process of deliberation will not enhance democratic outcomes that do not reflect complete agreement. Drawing upon Miller (1992), it simply means that the transformation might reflect a mechanism less oppressive than that which results in a conclusion favoured by all participants. This second-generation articulation results in what a significant portion of the literature categorises as the construction of 'public spirited' (Chambers 2003; p. 318) or 'other-regarding' preferences.

The most common way in which the notion of other-regarding preferences (often called 'social preferences') has been understood in a decision theory context emerges from Smith's concept of 'fellow feeling'. In this conception, the different levels of utility

that outcomes give other individuals are taken into account in the decision making process of a given agent. Essentially, you might say that one person has a preference for making sure another person's preferences are satisfied. More formally, if each individual's preference ordering can be represented by a function denoted as  $u_i$ , which is dependent on the choice from a bundle of goods  $X_i$ ; then a preference ordering motivated merely by individual concerns is denoted by the following  $u_i = u_i(X_i)$ . But when an altruistic component is substituted in, the following interdependent utility function is obtained  $u_i = u_i(X_i) + u_j(X_j)$ . In this case the utility of person 'j' is an element in the utility (and therefore preference ordering) of person 'i'. This is what is meant by rational choice theorists when they use the term rational altruism (Bardsley and Sugden 2006) or pro-social preferences (Van Lange 1999)

The first area I want to discuss, then, is how these other-regarding preferences, revised during deliberation, enhance democratic outcomes. And from this, I then want to briefly outline some of the empirical work taking place on this basis. To provide some structure to this question, it is possible to differentiate between the two cases that are often advanced simultaneously. The first involves the extent to which deliberative preferences are likely to reflect more agreement over a decision, and the second the extent to which they might help solve the stability argument presented by Black (1948) Arrow (1951), and Riker (1982).<sup>25</sup>

- i. Deliberative preferences lead to less disagreement
- ii. Deliberative preferences lead to more stable collective decisions

The concept of 'disagreement', then, is a popular issue with deliberative democrats, who place great importance on how it might impact the discursive process (Benhabib 1996a). A difference in language or mental capacity, for example, might make the progression of reasoned deliberation impossible. But within the confines of this chapter, the interpretation of difference that matters is how deliberation might change preferences to create outcomes that are more acceptable to more individuals. Let me develop the case for why this might be seen as beneficial then.

To do this, I want to draw attention to Wollheim's (1962) paradox of democracy, which

---

<sup>25</sup> Dryzek and List (2003) also discuss how deliberation might reduce the propensity of individuals to behave strategically in voting.



deals with the theoretical inconsistency of why a given individual will obey a collective decision at odds with their initial preferences. Putting aside a discussion of his various solutions, the salient question is, of course, on an aggregate level, why is Wollheim's paradox a problem? Does it matter if 51% of the population imposes a decision on the other 49%? Is a collective decision with a larger majority in some sense better? Intuitively, this comes down to the question of Beetham's (1991) second aspect of legitimacy, justifiability, which dictates that legitimacy is dependant on (a) the degree to which the collective decision matches the substantive goals of the people (judged on a sliding scale from minority, through plurality, majority and at the extreme unanimity), and (b) the degree to which it leads to normatively desirable ends (Parkinson 2003). The deliberative claim, to put it quite simply, is that preference revision towards more agreement (but not necessarily consensus) therefore leads to more legitimate democratic outcomes.

Due in large part to the dominance of an aggregation mechanism that requires individuals to simply tick a box stating only their most preferred outcome, this has been the traditional intersection between democracy and difference. It is though, only half of the story. Because deliberative models of democracy allow individuals to support their case with argumentation, the intensity of preference, implied by the phrase 'more acceptable', also becomes an important factor to consider. A given democratic outcome might very well enjoy the support of a majority of individuals. But if a particular minority feels much more strongly in opposition, then an argument can be made that it might not entirely satisfy the maxim of justifiability. Again, the notion of other-regarding implies that whilst an individual might not change their preferences entirely to match those of another person, they will at least take on the other individuals' perspective; seeing things from their side and ultimately revising the strength of the ordering. It produces what Mansbridge et al (2010; p. 78) refer to as a 'change in the strength and conviction' with which preferences are held.

To understand exactly how this works, let me provide an example that draws on both decision and social choice theories. To make the comparison with deliberative theory simple, individuals are argued to enter the deliberation with purely self-interested preferences. The process is then assumed to revise their desires, resulting in an ordering which can be depicted by a second utility function similar in structure to the one discussed above. For example, say we have three agents denoted by J, K and L. After a successful

deliberation over a particular issue, each comes to the opinion that their final preferences should now be based on a combination of both self-interest and a regard for others. Their new, revised preference function will therefore be dependent on a combination of the elements  $u_J(X_J)$ ,  $u_K(X_K)$ , and  $u_L(X_L)$ . Linking this with deliberative theory, if every individual in the hypothetical were to then value each other's preferences identically to their own, then they would all be choosing according to what Harsanyi (1955) famously referred to as an 'ethical utility function'.

But the very fact that most deliberative democrats choose to weaken the normative claim to the phrase other-regarding implies that whilst important, the ethical element does not entirely replace that of self-interest. What this translates to is some form of weighting between each component, which allows varying degrees of pro-sociality to exist in different preference orderings. Van Lange (1999; p.338) terms this a 'pro-social orientation'. To see how this translates into practice, consider the following example with J, K and L, who have the following cardinal preferences over two possible collective outcomes  $x$  and  $y$ .

Table 2.1: Individual preferences over two alternatives pre deliberation

Individual	Preference Function	x	y	Ordering
J	$v_J = u(X_J)$	100	80	$u_J(x) > u_J(y)$
K	$v_K = u(X_K)$	100	60	$u_K(x) > u_K(y)$
L	$v_L = u(X_L)$	10	100	$u_L(x) < u_L(y)$
Aggregate	$\bar{u}(X)$	70	80	N/A

Graphically, this data can also be plotted on a chart measuring  $u(x)$  on one axis,  $u(y)$  on the other, and with a  $45^\circ$  line providing what effectively denotes indifference between the two possibilities. Intuitively then, the shaded blue area that lies between the three individual's preferences, is representative of the conception of difference discussed earlier; it corresponds directly to the amount of disagreement between the citizens over the outcome they most prefer. Additionally, the shaded blue area that runs from L to the line of indifference is also indicative of the amount of disagreement with the aggregate choice that would obtain if a simple pair-wise comparison were to be made.

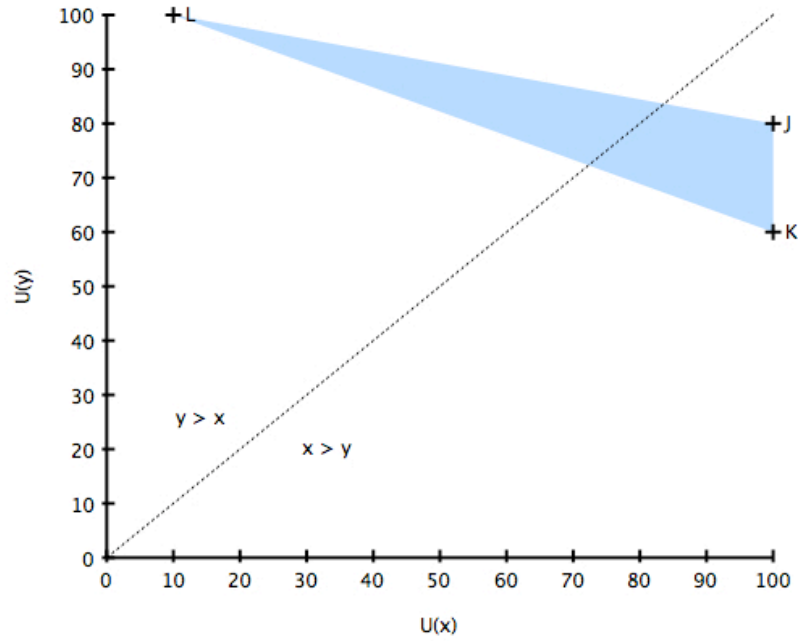


Figure 2.1: Preferences pre deliberation

After the process of deliberation has taken place, preference functions of all agents can then be modelled by the following expression:

$$v_i = (\alpha)u_i + (\beta)\bar{u} \quad \alpha + \beta = 1$$

For convenience, now imagine that we take the value of  $\alpha$  to be 0.3, and  $\beta$  to be 0.7, which roughly translates to the proposition that each individual now values societies' preferences to be of double importance to their own. A little calculation then gives the following other-regarding cardinal utility values for the policy outcomes  $x$  and  $y$ :

Table 2.2: Individual preferences over two alternatives post deliberation

Individual	Preference Function	$x$	$y$	Ordering
J	$v_J = \alpha[u(X_J)] + \beta[\bar{u}(X)]$	79	80	$u_J(y) > u_J(x)$
K	$v_K = \alpha[u(X_K)] + \beta[\bar{u}(X)]$	79	74	$u_K(x) > u_K(y)$
L	$v_L = \alpha[u(X_L)] + \beta[\bar{u}(X)]$	52	86	$u_L(x) < u_L(y)$

In this example, the most obvious effect that the addition of an ethical element to each preference function has, is to flip the preference ordering of individual J. But whilst this particular result is certainly sufficient, is not however necessary to make the claim that deliberation reduces disagreement. What should become apparent, and this is where a

second graphical representation is useful, is that even though agents K and L do not mimic this behaviour by changing their preferences entirely, they do however adjust the intensity of the ordering. The revision process therefore makes preferences tend towards a point of unanimity, but because the ordering retains a partially self-interested motivation, there is still some degree of difference between each agents' preference ordering. There is convergence, but not necessarily at the level of expressed preferences, which is displayed nicely by the reduction in size of the shaded area.

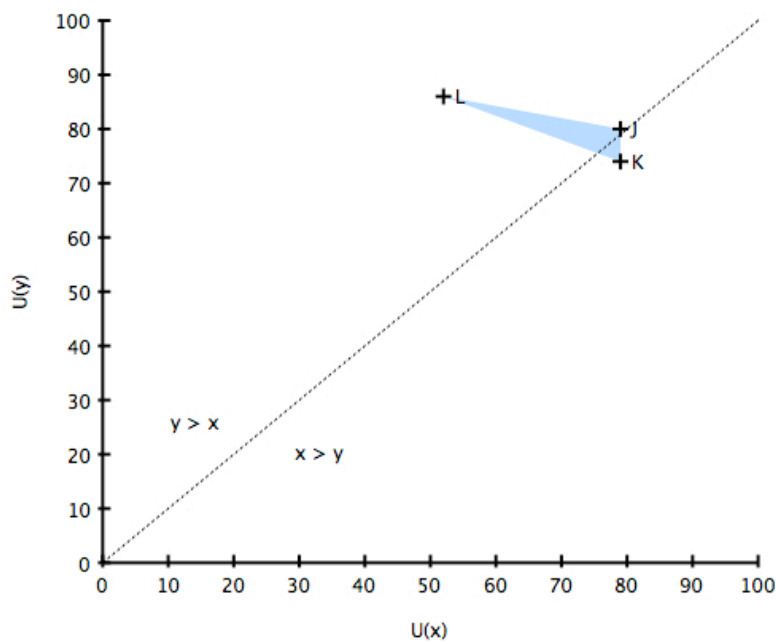


Figure 2.2: Preferences post deliberation

I now want to consider the second argument upon which other-regarding preferences, formed through deliberation, are claimed to secure better collective decisions. The concept of stability is most thoroughly explored in the social choice literature, popularised particularly by the US based Rochester School. Founded upon Arrow's (1951) now infamous possibility theorem, Riker (1988) took the result and offered a logically derived argument that all democratic outcomes were meaningless, or due to the inability to guarantee stability. Drawing upon Condorcet cycles (1785), Riker showed that transitive social preference orderings, whereby individuals' preferences are aggregated in an attempt to uncover what 'society' wants, could not be guaranteed without one of four weak axioms being broken: universal domain (U)<sup>26</sup>, Pareto

<sup>26</sup> Which states that the aggregation mechanism should be capable of taking as its domain of operation all logically possible orderings by individuals, or in other words, any pair of preferences should be able to be compared.

inclusiveness (P)<sup>27</sup>, independence of irrelevant alternatives (I)<sup>28</sup> and non-dictatorship (D)<sup>29</sup>. The inference from which was the potent result that democracy might yield an outcome where, for example, given three policy proposals,  $x$ ,  $y$  and  $z$ , the following social ordering might obtain:

$$r = x > y > z > x$$

Accordingly, Riker used this theoretical possibility to critique the entire notion of democracy, and, persuaded a large portion of the discipline of political science ‘to doubt that the content of social welfare, or public interest, can ever be discovered by amalgamating individual judgments’ (1982; p. 137). But his leap from a theoretical possibility to the certainty of casting a judgment upon all democratic outcomes is at best rather over-enthusiastic, and at worst wholly cavalier. Simply because cycles cannot be prevented without violating a logical axiom does not mean that they are guaranteed. Further still, it is not clear whether all of (U), (P), (I) and (D) are as applicable to deliberative models of decision-making as they are to those formed on purely aggregative principles. Indeed, whilst it seems appropriate that any logical combination of preferences should be allowed into the deliberative phase (satisfying universal domain), preference revision explicitly attempts to prevent certain orderings from being taken out of it. The aggregation stage of the decision-making process is therefore characterized by the input of other-regarding preferences, and because, by their very nature, commonality now exists between how individuals rank sets of alternative outcomes, the possibility of cycles appearing is greatly diminished.

For example, an intransitive social ordering can be produced when three individuals have the following self-interested cardinal preferences over three possible alternatives  $x$ ,  $y$  and  $z$ :

---

<sup>27</sup> Is a weaker version of positive responsiveness, and guarantees that if all individuals prefer  $x$  over  $y$ , then the social choice will also prefer  $x$  over  $y$ .

<sup>28</sup> Which states that a social choice between  $x$  and  $y$  will only depend on how individuals rank  $x$  compared to  $y$  in their personal preference orderings.

<sup>29</sup> Is a weaker version of anonymity, and guarantees that no named individual should be able to determine a social choice in all circumstances (in the sense that the social choice coincides with the ordering of that individual whatever others may think).

Table 2.3: Individual preferences over three alternatives pre deliberation

Individual	Preference Function	x	y	z	Preference Order
J	$v_J = u(X_J)$	100	90	75	$u_J(x) > u_J(y) > u_J(z)$
K	$v_K = u(X_K)$	50	100	95	$u_J(y) > u_J(z) > u_J(x)$
L	$v_L = u(X_L)$	60	20	100	$u_J(z) > u_J(x) > u_J(y)$
Aggregate	$\bar{u}(X)$	70	70	90	

By pair wise comparison, then, the following intransitive ordering obtains:

$$r = x > y > z > x$$

However, following the same process as above, whereby the consequence of a deliberation is to shift each individual onto an other-regarding preference function of the following form:

$$v_i = (\alpha)u_i + (\beta)\bar{u} \quad \alpha + \beta = 1$$

And where for consistency,  $\alpha = 0.3$  and  $\beta = 0.7$ , then the following preferences are constructed:

Table 2.4: Individual preferences over three alternatives post deliberation

Individual	Preference Function	x	y	z	Preference Order
J	$v_J = \alpha[u(X_J)] + \beta[\bar{u}(X)]$	79	76	85.5	$u_J(z) > u_J(x) > u_J(y)$
K	$v_K = \alpha[u(X_K)] + \beta[\bar{u}(X)]$	64	79	91.5	$u_J(z) > u_J(y) > u_J(x)$
L	$v_L = \alpha[u(X_L)] + \beta[\bar{u}(X)]$	67	55	93	$u_J(z) > u_J(x) > u_J(y)$

Which in turn produces the following transitive social ordering:

$$r = z > x > y$$

An altruistic element in each individual's utility function thus has the consequence of creating a shared way of ranking outcomes between different individuals. This link reduces the level of disagreement amongst the three individuals, resulting in the phenomenon of single peakedness (Black 1948, Dryzek and List 2003). This, in turn,

prevents the possibility of a Condorcet cycle from appearing. Deliberative preference revision, quite forcefully in theoretical terms, hence provides an effective counter-argument to Riker's position: democracy can still be meaningful in the sense the outcome reflects the will of the people.

Taken together, these examples demonstrate the usefulness of a rational choice analysis of the claim that deliberation leads to more public-spirited preferences. In combining decision and social choice theories, it is possible to conceptually unpack firstly what deliberative theory explicitly means, and secondly, how this translates to democratic decision-making that justifies the revision as beneficial. As I suggested in the introduction, though, it is also imperative for deliberative theorists to engage with empirical reality. Just as first generation articulations of deliberation and deliberative democracy were confronted with the fact that consensus is unlikely and even impossible, it is equally the case that the weaker argument of public spirited/other-regarding preferences needs to be investigated.

The decision theoretic interpretation of the revision process as one that simply adds an altruistic element into each individual's utility function does appear quite reconcilable with deliberative theory. It elucidates elegantly what the phrase 'other-regarding' means, and then matches up nicely with two key arguments behind why deliberative preferences might be considered superior to those formed solely upon self-interest. However, this is just *one* possible interpretation of the concept. Altruism, whilst certainly convenient and evidently theoretically successful, is not the only way in which the concept of other-regarding preferences might be understood. Indeed, within the scope of decision theory, a number of other potential conceptualisations have been generated (see Fehr and Schmidt 2006; pp. 636-644). In turn, each with their own degrees of success and failure in securing the democratic benefits that normative political theorists claim deliberation obtains. The phrase other-regarding, for example, can also quite plausibly be used as a proxy for the notion of reciprocity, where individuals act with altruistic preferences only when they expect others to do the same. Or indeed, it is equally plausible that the structure of deliberative preferences might mirror something akin to inequality aversion, where an individual favours outcomes that offer the same amount of utility to every member of society. Each of these also seems a plausible interpretation of the concept (Buchan et al 2006).

From an empirical standpoint, the challenge for deliberative theorists is therefore to try and map the structure of other-regarding preferences that deliberation yields. In doing so, different conclusions can be drawn and deliberative models of democracy adjusted accordingly. On the level of preference revision, there is a small but developing literature in political science that has begun to look at this type of question (Delli Carpini et al 2004; Fishkin and Luskin 2005; Mendleberg 2002). Niemeyer (2004), for instance, has utilised Q methodology and found evidence for the claim that deliberation induces more environmentally friendly preferences<sup>30</sup>. Luskin et al (1999, 2000, 2004) have also found support for the claim the deliberation causes preference revision, although different experiments have resulted in different ‘types’ of shift. Finally in another study, Farrar et al (2010) use the results from a deliberative poll to test the hypothesis regarding single peaked preferences. In doing so, they find some supporting evidence for the theoretical proposition over topics they term less salient. As they put it themselves: ‘deliberation tends to produce net attitude change and bring preferences closer to single-peakedness, at least on issues of low to moderate salience’ (Farrar et al 2010; p. 344). The problem for these experiments, however, is the difficulty in coming up with a suitable methodology that can distinguish between different other-regarding preference structures.<sup>31</sup> Even Farrar et al (2010) who proceed on a social choice informed basis, do not really engage with the decision-theory side of the analysis. On the other side of the coin, studies completed within economics such as Dawes et al (1977), Orbell et al (1988) and Roth (1995), do pay closer attention to the structure of preferences, but then do not really test the impact of deliberation as conceptualised in chapter one. To put it bluntly, I would contend that whilst preference revision has begun to enjoy some significant empirical attention, there is much work that remains to be done. And in particular, a more technical approach, such as decision-theory, is required to really understand the revisionary effect of deliberation.

Let me briefly summarise this section then. At the level of preferences, deliberation is argued to shift individuals from one utility function characterised predominantly by self-interest, to another, which reflects a more pro-social perspective. When presented with possible courses of action in democratic decision-making, an individual is therefore modelled to go through the following process of reasoning:

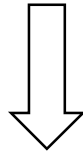
---

<sup>30</sup> Niemeyer (2004) is also an example of deliberation’s effect on issue revision, something I will consider briefly in the following section.

<sup>31</sup> See, for example, Charness and Rabin (2002), Charness and Ertan (2002), Cox (2004) or Kagel and Wolfe (2001) for some experimental economic approaches to this question.



What policy alternative do I favour, given my self-interested preferences?



Deliberative revision

What policy alternative do I favour, given my other-regarding preferences?

However, deliberative theory seems to suggest something more than just creating individuals whose preferences take into account the impact of a decision on others. It also seems to suggest a change in the way these preferences are used by the individuals in question. I will take up this matter specifically in section 2.4 regarding the question of agency revision.

### **2.3 DELIBERATION AND ISSUE REVISION**

---

But before I progress to the most crucial section of this chapter, to complete the analytical framework I want to very briefly say something about how deliberation might positively affect the process in terms of the issue that is being considered. When democratic theory is understood in purely aggregative terms, the issue at stake is often taken as exogenous to the process. Individuals maximise their utility by selecting the course of action most likely to satisfy their preferences over the set of alternatives on offer. In many cases though, particularly regarding the election of representatives but also in referenda, sets of alternative policies are bundled together in groups, or sets of mutually exclusive outcomes. In particular, this is often the case where ideologies, rather than policy effects, are the true focus of the collective decision. In the UK, for example, throughout the 1980s and 1990s the political system was dominated by the ideological debate between privatisation and public ownership. Each party attempted to galvanize supporters on these general principles, and then put forward policy proposals (in the form of candidates) for the electorate to vote on. The question, then, is why might deliberation be beneficial in situations like this?

Let me illustrate my argument by way of a simple formal example. Consider a collective decision that is put before an entirely aggregative democratic process. The

decision, however, actually contains two issues, each with two policy alternatives such that:

First policy decision:  $w$  or  $x$   
 Second policy decision:  $y$  or  $z$

Now assume for some reason (i.e. ideological or even for manipulative purposes) that  $w$  and  $y$  are offered to the electorate as linked, such that they are represented by the profile of policies  $\alpha$ , and so too are  $x$  and  $z$  denoted by  $\beta$ . Say we have three individuals J, K and L, and they have the following cardinal preferences:

Individual	$u(w)$	$u(x)$	$u(y)$	$u(z)$	$u(\alpha)$	$u(\beta)$
J	80	20	30	70	110	90
K	90	10	20	80	110	90
L	100	0	10	90	110	90

Clearly then, the following is true for all three individuals:

$$u(\alpha) > u(\beta)$$

Which means when put to a democratic vote, the profile of policies in  $\alpha$ , namely  $w$  and  $y$ , will obtain. However, whilst for all three people it is the case that:

$$u(w) > u(x)$$

It is also true that:

$$u(z) > u(y)$$

The bundling of two issues into a single collective decision, whilst still allowing the choice of a most preferred outcome to any of its alternatives on offer, in this case clearly subverts the true preferences of the individuals involved. Two ways in which discussion and contestation of the issue at stake can have an impact then, is either to suggest entirely new policy proposals that may reflect even more closely the preferences

of the people; or in this case, deliberation can be used to decide specifically *what* the issue at stake actually is. In this case, a deliberation instituted before aggregation might identify that the first and second policy decisions are both quite controversial, in the sense that individuals in the model prefer  $w$  to  $x$ , and  $z$  to  $y$  by large margins. By revising the issue, however, specifically by identifying and unpacking the two different decisions to be made within it, and placing them into separate contexts, the democratic process is then able to accommodate the profile of choices  $w$  and  $z$ , represented by  $\theta$ , and  $x$  and  $y$ , by  $\omega$ :

Individual	$u(\alpha)$	$u(\beta)$	$u(\theta)$	$u(\omega)$
J	110	90	150	50
K	110	90	160	30
L	110	90	190	10

Given that the following is true for all three individuals:

$$u(\theta) > u(\alpha) > u(\beta) > u(\omega)$$

Deliberative issue revision therefore allows the profile  $\theta$  to be chosen. Since it turns out to be more preferable for all three individuals than  $\alpha$ , the collective decision that is reached post-deliberation can thus be seen as better. In the same way in which less disagreement performed better on Beetham’s (1991) legitimacy framework, it can be argued that the collective decision in this example more closely satisfies the will of the people. Empirically then, the question of issue revision is linked strongly to that of preference revision. Niemeyer’s (2004) use of Q methodology, for instance, allows for both to be investigated simultaneously, although it consequently makes distinguishing between the two phenomena impossible. Again, then, the topic of issue revision has begun to enjoy some empirical attention, and is certainly an avenue that requires further investigation and analysis.

## 2.4 DELIBERATION AND AGENCY REVISION

---

This leads me to the final level at which deliberative revision is possible, and concerns the way preferences over different outcomes translate into practice. By this, I refer to

the mode of reasoning that combines the set of possible alternatives with preferences, and then determines an individual's course of action. In this sense, it applies much more at the level of the individual, rather than the nature of the democratic decision which social choice theory is concerned with. From the perspective of decision theory, it relates explicitly to the concept of rationality, and more specifically, how things like probability and uncertainty are factored into the calculation process of a given individual. More often than not then, decision theory is taken as identical to either von-Neumann Morgenstern's (1944) expected utility theory (EUT), or Savage's (1954) subjective expected utility theory (SEUT). Individuals are modelled to act on their preferences by comparing the expected utility that each outcome will secure. For both EUT and SEUT, this figure is obtained simply by multiplying the (subjective) probability of their preferences being satisfied given their action, by the utility that it yields. For example, given two possible outcomes where  $x$  is preferred to  $y$ , then a given individual will only act in accordance to bring  $x$  about if, and only if:

$$p[u(x)] > p[u(y)]$$

One obvious way in which revision can interact with this articulation of decision theory, is the value of  $p$  that is assigned to each outcome. As the given parameter fluctuates, then so too does the prescribed rational course of action that individuals will pursue. It should be noted that this is something very different to the notion of preference revision, since it is not the values of  $u(x)$  or  $u(y)$  that are altering. Rather, it is something in the structure of the decision calculus, or reasoning, that has changed. However, as logical as the assumptions behind EUT and SEUT are as rational decision rules, a number of questions to the validity of the model in certain salient cases have been raised. One of the most famous of which, for example, goes back to the experimental work of Kahneman and Tversky (1981), who asked a number of subjects to make a decision given the following information and options:<sup>32</sup>

- i. Each person starts with 1000 [Israeli pounds]
- ii. Now choose between:  $x$  = certain gain of 500, and  $y$  = 0.5 chance of gaining 1000

---

<sup>32</sup> For simplicity, monetary payoffs are treated as equivalent to utility payoffs.

In this first case, 84% of agents stated that they would take the course of action  $x$  over  $y$ . A second group was then given the following slightly different information:

- i. Each person starts with 2000 [Israeli pounds]
- ii. Now choose between:  $x'$  = certain loss of 500, and  $y'$  = 0.5 chance of losing 1000

This time, 69% of people chose  $y'$  over  $x'$ , even though from the perspective of EUT (or SEUT),  $x$  is mathematically identical to  $x'$ , as are  $y$  and  $y'$ . Kahneman and Tversky's result, then, has largely been used to argue for the importance of framing in a rational choice approach to decision-making, with two avenues of exploration specifically emerging as a result. The first is to jettison some of the standard assumptions behind EUT in order to come up with a different model of decision theory applicable in these circumstances. Kahneman and Tversky's case of framing requires a theory of utility maximisation that takes into account a differentiation that some individuals make between potential 'losses' and 'gains' in utility. Prospect theory (Kahneman and Tversky 1981), as an example of a non-EUT model of decision theory, was therefore offered as an alternative approach. It takes into account 'reference based' information, and can therefore distinguish between gaining and losing utility. Loomes and Sugden (1982) on the other hand, have used behaviour that departs from EUT and SEUT to generate a model of decision theory they call 'regret theory'. This model predicts that individuals will choose on the basis of minimising the potential regret they might feel in decision-making under uncertainty.

In a useful review article, Starmer (2000) takes the position that the eventual goal of decision theory is to come up with a single formal model of choice that can be used to explain all human behaviour. Whilst this endeavour is surely worthwhile, a different perspective might also be taken. Instead of generating a universal model of choice, an alternative line of enquiry is to look at the circumstances that provoke certain models to be used by individuals. Under this interpretation, the claim is that in particular cases, individuals will be more or less likely to rely on different models of agency. A plausible case, for instance, might be made that the prioritisation of objectivity and rationality within deliberation could induce individuals to act in accordance with the most simple, and objective model of agency: EUT. Work done by Loomes, Starmer and Sugden (2003), although not strictly concerning deliberation, adds at least some potential

support for this idea in their investigation of the erosion of reference dependent choices (indicative of prospect theory) during repeated market interactions. Since deliberation is also an example of a dynamic learning process, an intuitive claim to test along these lines would therefore be whether allowing a reasoned deliberation in Kahneman and Tversky's (1981) example would affect the decisions made by the individuals in question. Would it induce individuals to employ EUT over that of prospect theory? This presents one possible hypothesis to explore on the level of agency revision.

However, in this thesis I want to consider a more fundamental change in which the concept of agency revision might be understood. Decision theoretic models like EUT, SEUT, prospect theory and regret theory all share the assumption that reasoning on behalf of preferences takes place at the level of the individual. They are premised upon methodological individualism (Weber 1968 [1922]). Preferences might include a consideration for another agent, but the individual is still modelled to go through the following reasoning mechanism:

- Given the alternatives on offer and my preference ordering over them, what course of action should I take?

In the following section I want to explore how the individualistic approach to decision theory might be relaxed, and in particular how a model of 'group agency' might be applied to deliberative theory.

### 2.5.1 DELIBERATIVE AGENCY REVISION: TEAM REASONING

In the section on preference revision, I suggested that deliberative theory was arguing more than simply that deliberation transformed preferences to an other-regarding nature. To see what I mean, consider once more Wollheim's paradox of democracy. In an attempt to establish why individuals freely and legitimately surrender their preferences to those produced by the democratic process, Wollheim explored a number of potential solutions<sup>33</sup> before settling on an answer he felt truly explained what was happening:

By distinguishing between the terms 'voter' and 'democrat', and that whilst

---

<sup>33</sup> The first was by denying that the citizen was committed to the belief that his original preference for A, in the face of the social choice for B, was correct; the second by denying that the citizen actually believes the democratic social choice is correct, i.e. remains committed to a belief in policy A, and the other by simply consigning it 'to the flames' (ignoring the issue).

policy A might be the choice of an individual acting merely as a voter, if a social choice results in policy B, then this becomes the preference of an individual acting as a true democrat.

(Wollheim 1962)

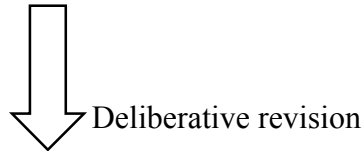
One aspect of the interpretation of this distinction between the terms ‘voter’ and ‘democrat’ is covered by the notion of preference revision. The preference ordering of a voter is represented by a given initial utility function; and that of the democrat, by a function representative merely of the resultant social choice (an approach similar to Margolis 1981, 1984). An individual, after participating in the aggregative democratic process, simply shifts from one ordering to another due to any of Elster’s (1983) proposals. But rather critically, this mechanism does not seem to entirely capture what is truly meant in Wollheim’s suggestion. The term democrat implies more than a particular preference ordering. What I suggest is missing, is due to the reliance on individualism that both EUT and non-EUT models of agency have. For individuals to label themselves as democrats, they are clearly embracing an identity beyond that represented simply by their utility functions. They are defining themselves with reference to their membership of a particular group of individuals. In terms of deliberative democratic theory, Barber (1984; p. 200) sums this up quite nicely when he claims ‘in place of I want Y, the strong democrat must say Y will be good for us’. Or, for another example, Thompson (2008) states:

Discussions framed by asking participants, “what action should we, as a group, take?” come closer to the deliberative democracy (creating a genuinely public opinion) that they favour [...]

(Thompson 2008; p. 503)

Deliberation is expected to not only change the preferences of an individual, but also to transform the way they then reason upon them. Or to put it more strongly, it assumes a very specific case of agency revision takes place. Instead of asking ‘how should I act’, deliberation is claimed to make individuals ask themselves the question ‘how should we act’. This distinction between pronouns is crucial as it reflects a departure from the wholly individualistic approach offered in all the previous different models of agency. In the language of rational choice theory, revision due to deliberation is argued to produce ‘team reasoning’ amongst individuals.

What course of action should I take, given my preferences?



What course of action should we take, given our preferences?

The idea that teams of individuals can count as agents in their own right, then, is nothing new. Indeed, it has been proposed in various different guises by, amongst others, Hodgson (1967), Regan (1980), Gilbert (1987, 1989), Hurley (1989), Sugden (1993, 2000, 2003), Bacharach (1999, 2006) and Coleman and Rose (2008).<sup>34</sup> To see how it works, I first want to outline an argument that gives weight to the viability of collective reasoning as a model of agency. More specifically, consider a particular type of encounter between two individuals known as the ‘Hi-Lo’ game:

		Individual $j$	
		high	low
Individual $i$	high	10 , 10	0 , 0
	low	0 , 0	1 , 1

Figure 2.3 The Hi-Lo Game

In this interaction, two individuals  $i$  and  $j$  are required to choose one of two strategies, either ‘high’ or ‘low’. It is clear that the interests of the players are intrinsically tied together: they achieve a preferred outcome only when they co-ordinate their action. From a standard game theoretic analysis, two Nash equilibria obtain. Both [high, high] and [low, low] are considered equally rational from an entirely self-interested perspective. Classical individualistic decision theory cannot explain why, when tested empirically, high is played with a probability almost equal to one (Sugden 1995).<sup>35</sup> What makes [high, high] the seemingly more rational course of action for both individuals to take than [low, low] is the implicit appeal to the principle of payoff dominance, which almost all individuals identify when presented with this decision. Team reasoning can incorporate such a consideration, since it takes the following

<sup>34</sup> A related literature on collective intentions is also present, see Searle (1990) or Bratman (1993)

<sup>35</sup> One possible suggestion is that other-regarding preferences alone may provide a rational basis for choosing the high strategy. I will demonstrate how this logic does not apply to the Hi-Lo game, or indeed particular types of prisoners’ dilemma games, in chapter four (section 4.3).



approach, expressed most simply in propositional logic by Gold and Sugden (2008; p. 289):

- i. I am a member of the group  $S$
- ii. It is common knowledge in  $S$  that each group identifies with  $S$
- iii. It is common knowledge in  $S$  that each member of  $S$  wants the value of  $U$  to be maximised
- iv. It is common knowledge in  $S$  that the course of action  $A$  uniquely maximises  $U$

I should choose my component of  $A$

Applying this to a high-low game, the two individuals taken together form the group  $S$ . The maximum value of  $U$  is given by summing payoffs in each box, in this case it equates to  $10 + 10 = 20$ . And finally, the course of action  $A$ , denoted by the strategy that achieves this, is for both individuals to play high. In this sense, the Hi-Lo game provides an extremely clear example and indeed argument for the validity, of team reasoning in action. It is the most logical form of reasoning that truly explains why an individual, acting rationally, will choose the ‘high’ strategy.

Within decision theory, the discussion of the process that causes<sup>36</sup> an individual to team reason has largely been dominated by two similar, but subtly quite different conceptualisations. The first, offered by Sugden (1993, 2000, 2003), requires slightly more than agency revision, and relies on the assumption that team preferences can exist in the same manner and form as those represented by an individual’s utility function. He terms his approach as ‘team directed reasoning’ (Sugden 2000; p. 1182). Decision theory therefore remains a matter of utility maximisation, with the preferences of the team rather than the individual being satisfied. But interesting, Sugden’s approach also pays particular attention to the notion of assurance, and whether an individual has reason to believe that everyone else in a given team is likely to reason in an identical manner. In this way, whilst individual and team preferences might take the same form, there is a slight difference in the reasoning mechanism that links them to individual behaviour. In cases of their own preferences, individuals need only consider what

---

<sup>36</sup> Whilst this point of departure is certainly important, I want to stress that both conceptualisations still result in the same process: i.e. individuals asking themselves the question ‘what should we do’.

single course of action is most likely to satisfy them. In cases of team preferences, though, each individual is then required to choose an array of different courses of action: one for each team member, and they select their own from this list. When this is combined with assurance, i.e. reason to believe that each other individual will ‘play their part’ then team-directed reasoning is employed. For simplicity then, it might be represented by the following propositional account (Gold and Sugden 2008; p. 303):

- i. I am a member of S
- ii. I identify with S and acknowledge U as its objective
- iii. In S, there is cross-personal common reason to believe that each member of S identifies with S, and acknowledges U as the objective of S
- iv. In S, there is cross-personal common reason to believe that A uniquely maximises U
- v. In S, there is cross-personal reason to believe that each member of S endorses and acts on mutually assured team reasoning

I should choose my component of A

The other approach to collective reasoning is that offered by Bacharach (1999, 2006). His account envisages the process of team reasoning within variable frame theory, and does not claim that groups themselves can have preferences. Rather, he argues that all goals must also be representative of the preferences of at least one agent in the overall process. From these two points of departure, his model then illustrates that individuals, depending on the frame they inhabit, will view interactions from either an ‘I’ or a ‘we’ perspective. As such, the unit of agency is not chosen by the individual per se, but rather primed by the presence of environment they inhabit. This environment, crucially, is exogenous to the model of decision-making, and is depicted as a cognitive context that acts upon the individual in question. In cases where the ‘we’ frame is more prominent, individuals will team reason. In cases where the ‘I’ frame is more obvious, then individuals revert to classical decision-theory. Accordingly, his model might be represented more simply by the following slightly different set of propositions (Gold and Sugden 2008; p. 297):

- i. I am a member of T
- ii. It is T-conditional knowledge ( $\omega$ ) that each member of T identifies with S

- iii. It is T-conditional knowledge that each member of T wants the value of U to be maximised
- iv. It is T-conditional knowledge that P uniquely maximises U, given the actions of non-members of T

I should choose my component of P

Unlike Sugden's account, Bacharach's model explicitly demonstrates the rationality of team reasoning in the same way that decision theory does with individual action. As Sugden and Gold (2008; p. 296) make clear: 'for any given individual, if she identifies with S and wants U to be maximised, it is instrumentally rational for her to act as a member of the T, the team of like-minded individuals'. What matters in this model is the value of  $\omega$ . When it is sufficiently high for a given agent, they employ the team reasoning account of agency.

From the perspective of reconciling these accounts with deliberative theory then, both Bacharach and Sugden's model provide a coherent empirical comparison. Crucially, both envisage the factors that promote team reasoning as coming from outside the rational choice framework. In Bacharach's language, it is a question of framing. Deliberation in a political context can clearly be viewed as an activity that directly contributes to this process in a decision-making situation. Discussion and communication, for example, are dynamic factors that cognitively impact upon an individual before a choice needs to be made. On this basis, in the following chapter I will consider the link between deliberation, framing and teams in much more depth, and in particular explore which features of deliberation are likely to prime the 'we' frame. But before I do that, I want to say three things about *why* this should be considered a normatively appealing direction for revision in the first place.

### 2.5.2 THE EFFECTS OF TEAM REASONING

Having made the claim that deliberative democratic theory often assumes a particular type of revision at the level of agency, I want to consider the question of the positive effects on a political system of creating team reasoners. To answer this, I will split my analysis and argument into three related sections:

- i. Team reasoning helps to 'solve' social dilemmas

- ii. Team reasoning provides a ‘bond for society’ (Hollis 1998)
- iii. Team reasoning contributes directly to ‘community generation’ (Cooke 2000)

Firstly then, consider the impact that team reasoning has at an individual level. When presented with social dilemma type games, for example the Hi Low encounter, individualistic reasoning does not result in the most beneficial outcome for those involved. Indeed, this problem is further compounded in games like the Prisoners’ Dilemma, where the duality between the outcome that is best for a single individual and the outcome that is best for both, is clear. Take the following (non symmetric) game as an example then:

		Individual $j$	
		co operate	defect
Individual $i$	co operate	3 , 3	-2 , 4
	defect	4 , -3	-1 , -1

Figure 2.5 A Prisoners’ Dilemma Game

From a classical rational choice approach, game theory (the interactive variant of decision theory) prescribes that for both players, defect is the dominant strategy. For agents applying an individualistic mode of agency, they receive a better outcome if they choose defect no matter what strategy the other individual employs. In this case, the point [defect, defect] is therefore defined as a Nash equilibrium. However, this means that a clearly sub optimal position is arrived at. Both players prefer two other possible outcomes. Individual  $i$ , for example, is better off at [defect, co operate], whilst player  $j$  is better off at [co operate, defect]. And most crucially, both are better off at [co operate, co operate]. How, then, can individuals reason so they both play the co-operative strategy and end up at a more preferable outcome?<sup>37</sup>

<sup>37</sup> One answer that has been suggested is that social dilemma games of this particular structure might also be ‘solved’ by giving both individual players other-regarding preferences (Rabin 1993, Basu 2006). This is a point of contention I will consider in more depth during chapter four, where I devise an experimental setup that tests *only* for agency revision. For the purposes of my present argument, however, let me just state that this particular encounter is non-decomposable, which means co-operate cannot be sustained as a rational choice even on other-regarding preferences. Moreover, even if it could, it doesn’t diminish the argument that team reasoning also might also ‘solve’ the game by inducing a cooperative strategy choice.

If the game is viewed under Bacharach’s (1999, 2006) ‘we’ frame, then the strategic nature of the interaction is partly removed. Approaching the strategy choices using reasoning which asks “what should we do” makes it rational for both individuals to then select the choice of co-operate. Team reasoning leads to an outcome that is payoff dominant, or in other words, an outcome that represents the most preferred from a group’s perspective. In terms of political theory, social dilemmas such as the game above are examples of the problem of what Olson (1965) famously labeled ‘collective action’. Democratic participation such as voting, or action such as contributing money to charitable organisations, are all cases of decision-making where the outcomes for society and the individual are in conflict with each other. Individuals who employ team reasoning in these cases, therefore, are more likely to contribute or participate. To put it simply then, deliberative agency revision that creates more ‘team reasoners’ should be preferred if we judge such actions as voting and charitable donations to be valuable to society.

Building upon this, my next point extends this analysis and specifically considers what it means for two individuals to be engaged in team reasoning together. But whilst the first argument took a consequentialist approach in demonstrating the positive impact that team reasoning has on the outcome of the interaction, this second point is rooted much more in the nature of the reasoning itself. What does it mean to say that two individuals are engaged in team reasoning? In his final book *Trust within Reason*, Hollis (1998) focused explicitly on the concept of rationality in decision-making. Built around an allegory that describes a walk between two individuals Adam and Eve, Hollis demonstrates the problem of backwards induction that pervades individual models of agency. Let me quickly outline his case:

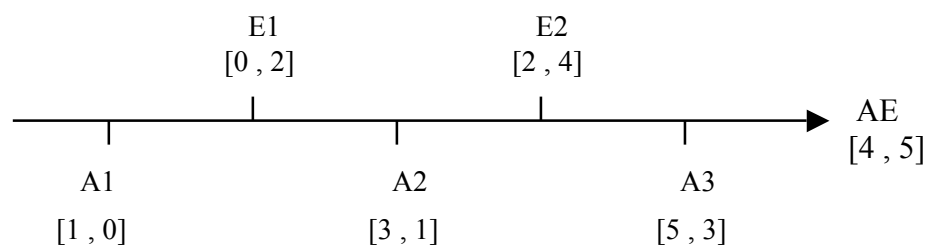


Figure 2.6: The Enlightenment Trail (Hollis 1998; p. 16)

Consider the two individuals Adam and Eve, who whilst walking down an imaginary road decide they would like to stop off at a pub. Along the trail are six different venues (A1, E1, A2, E2, A3 and AE), each providing different levels of utility as shown on the diagram above. Both Adam and Eve agree that as they progress along the walk, they will take it in turns to decide whether to stop off at that particular venue. Adam chooses at A1, Eve at E1, and so on. If all other pubs are passed, then AE will be the final stopping point for both. Individual reasoning, then, prescribes the following result. If both individuals choose not to stop at any of the first four pubs along the route, then Adam will clearly choose to stay at A3, since he prefers it to AE. Knowing this, however, Eve then has a rational incentive to stop at E2, one pub before, as she prefers this to A3. Yet this makes it rational for Adam to therefore choose to stop at A2, as he prefers this to E2. Again aware of this, Eve therefore would choose to stop at E1, which thus makes it rational for Adam to stop at A1, the first possible venue along the walk. In doing so, they end up at a venue that is clearly suboptimal: the outcome that is least preferred for Eve, and second least preferred for Adam.

As Hollis points out, both can do much better if Adam aims for the venue he prefers as second best. In this case, he might promise Eve that once they get to A3 he will decide to stop there. But under classical rational choice reasoning, Eve has no reason to believe Adam's promise. If she agrees to his proposal, Adam should take advantage of the situation, choose not to stop, and they both end up at *his* most preferred location of AE. Without any justification for Eve to believe Adam's promise, the backwards induction argument simply runs its course again, and they end up at A1. Hollis' (1998; pp. 137-142) solution, then, is to suggest that *trust* is included *within* the team-reasoning model of agency. If both individuals approach the walk from the standpoint of a group, and ask themselves the question 'what should we do' at each stage, then each has a justification to believe the promise made by the other individual. The very fact that an individual team reasons in this multiple stage interaction assumes they trust their co player to do the same, or else risk ending up at a less preferred outcome. As Hollis (1998; p. 159) puts it 'in seeing ourselves as persons with interdependent reasons for action, we clear the way for a liberal society as a community where trust is secured by mutual respect and generalised reciprocity among reasonable persons'. It is important to point out that this is a very different argument to the one presented in relation to social dilemma games. In that example, I made the case that team reasoning led to

certain *outcomes* that were beneficial. In this example the claim I make is that team reasoning itself *is* the better outcome.

Considered together, these two points form the thrust of my final argument. By demonstrating that team reasoning can be valued on both instrumental (outcome) and non-instrumental (processual) grounds, it is possible to combine them both into a broader point. Cooke (2000; p. 949), in discussing five different arguments for deliberative democracy, touches upon the claim that ‘the process of deliberation has a community generating power’. She draws out a number of possible features of this position, including Cohen’s (1996; p. 102) democratic assertion that it ‘expresses the equal membership of all in the sovereign body responsible for authorising the exercise of that power’; and the Habermassian ideal that deliberation requires all individuals to think about what counts as good reasons for all members of society. But, she also claims that the argument fails to really provide an adequate justification for deliberation in its own right. In particular, she claims that ‘the community generating-argument runs up against the problems [...] of how to show that deliberative participation in public affairs is *superior* (in its community generating effects) to non deliberative participation [...]’ (Cooke 2000; p. 950).

I would contend, then, that both the previous points can be used to answer this criticism. Let me start by assuming that team reasoning, as a product of deliberation, does result in the following beneficial outcomes:

- i. More participation in socially valuable, but individual costly, activity.
- ii. Higher levels of trust amongst individuals.

As Hollis (1998; pp. 150-154) argues, a point of comparison can be made with Rousseau’s ‘remarkable change in man’. The movement from individual to team reasoning represents a significant change in both the perspectives and actions of the individuals involved. Team reasoners trust others, and place the interests of society above that of the individual in a way that people employing individual agency do not. In doing so, their participation has a direct positive impact on the sense of community amongst a population.

To summarise the overall argument of this section then. If deliberation causes agency

revision to team reasoning amongst the individuals involved, which is then characterised by these two elements of community-generation; then deliberative models of democracy might then be justified alone on this basis. In short, there is a strong normative ground for the desirability of a political system that provides such a revisionary effect.

## 2.5 CONCLUDING REMARKS

---

In this second chapter, then, I have outlined a three-fold model of deliberative revision, drawing upon the rational choice approach to distinguish between the distinct levels at which this was possible. I then provided a discussion of the two most obvious levels, that of preferences and the issue at stake. In the final section of the chapter, I developed a case for the suggestion that deliberative theory argues for revision at another level that is often overlooked, that of agency. I then discussed a number of ways in which agency revision might be conceptualised, before focusing on the topic of ‘we’ thinking in the form of team reasoning.

After outlining Sugden (1993, 2000, 2003) and Bacharach’s approaches (1996, 2006), I settled upon the jointly held concept of framing as an appropriate line of further inquiry into the effects of deliberation. Following a similar sentiment expressed very briefly in Dryzek and List (2003; p. 12), I conclude this chapter with the following testable research question:

**Q** To what extent does deliberation make the ‘we’ frame more prominent and trigger agency revision, causing individuals to team-reason?

To answer this (fundamentally) empirical question, a number of preparatory stages must be considered. In the following chapter, I consider the conceptual link between framing, deliberation and team reasoning, which sets up the experimental approach outlined in chapters four and five.



## — CHAPTER 3 —

# FRAMING AND TEAM REASONING: HOW DELIBERATION PRIMES SOCIAL IDENTITY

### 3.1 INTRODUCTION

---

In the first chapter of this thesis, I discussed the notions of deliberation and deliberative democracy. In the second chapter, I then drew on the rational choice approach to analyse the various claims made regarding deliberation, and used it to identify three levels at which revision was possible. After noting that the third, that of agency, has been ignored both on conceptual and empirical grounds, I outlined three key arguments for a political system that triggered this type of revision. My task now, is to show how this might be reconciled with the concept of deliberation.

This chapter will address the following two points:

- i. A discussion of the theory behind, and empirical testing, of factors that have been found to induce individuals to identify as part of a team, therefore making them likely to employ team reasoning as a mode of reasoning.
- ii. An analysis of the specific features (and process) of deliberation to see how they reconcile with these determinants of social identity.

In demonstrating a theoretical argument for the link between deliberation, framing and team reasoning, it lays the foundation for the latter empirical investigation. To understand why individuals might *identify* as part of a team (and therefore team reason), a significant amount of conceptual path clearing is required though. For instance, what is meant by the term ‘identity’, what constitutes a ‘team’, and how do the two concepts interact with each other? Further still, to what extent can these terms even be applied in conjunction with the rational choice approach? Answering these initial questions should then provide a foundation for the construction of a model of identity formation – which can then be applied to the key features of micro-deliberative processes to theoretically test the normative claim that deliberation might induce team reasoning.

## 3.2 IDENTITY THEORY AND RATIONAL CHOICE: DEFINING THE ‘SELF’

---

### 3.2.1 DEFINING THE SELF: PERSONAL IDENTITY

Conceptual discussions of identity then, are predominantly located within the domain of social psychology. When any individual employs terms including ‘I’, ‘me’, or ‘you’, it is clear on one level that they are referring to a defined entity. But what is much harder to ascertain is the extent to which these labels are actually referring to something more nuanced, more complex than just a physical object. It is this distinction that led to the now established research agenda on the concept of ‘the self’.<sup>38</sup> To give a brief overview, then, the field can be subcategorised into two broad categories: the American, and the European traditions – a geographical distinction used primarily for its ability to offer a degree of memorability. I will deal with the latter in the next section.

The American tradition, then, deals with the self from a largely individualistic perspective – focusing on the person in question, and making reference to personal features that make an individual distinct from their contemporaries. For example, psychological research on this area has been dominated by investigation and discussion of terms such as self-schema, self-complexity, self-verification or self-affirmation. What should be immediately obvious is that all these terms are highly personalised, and are reflective of an approach focussed on the internal process of the individual during identity formation. An interpretation of the self-concept along these lines has led to experimental work predicated on the importance of the term after the hyphen then. Markus (1977), for instance, has looked at the extent to which individuals report the particular trait of independence as important to their self-definition. On the other hand, the idea of self-affirmation has been investigated by, for example, Koole, Smeets, Van Knippenberg and Dijksterhuis (1999) to show it reduces the prevalence of reflective thinking (ruminence) post failure. What remains central to these types of investigation though, is the assumption that identity is a highly personalised concept that can be studied without reference to a social context.

Personal identity is the individuated self – those characteristics that differentiate one individual from others within a given social context.

---

<sup>38</sup> The literature surrounding this concept is so vast, that Deaux (1992; p. 9) comments ‘in a recent edition of my social psychology textbook, for example, 11 self-hyphenated terms ranging from ‘self-concept’ to ‘self-serving attribution bias’ are defined in the glossary [...]’.

Relating this to the methodological approach taken by the previous chapter of this thesis, the concept of identity might seem a rather odd topic, particular as rational choice theory treats the notion with very little regard. Indeed, even this may be putting it a too strongly. Conceptually and operationally, decision theory depicts the self simply by the information contained in a utility function: individuals are defined only by what they desire. As such, stark similarities can therefore be drawn with the American tradition: identity is highly individualised and an entirely personal matter. To see this more clearly, it is helpful to consider the matter using a simple schema of reasoning:

- i.  $a$  is a defined personal identity
- ii. Preferring  $x$  to  $y$ , and  $y$  to  $z$  are characteristics of  $a$
- iii. Individual  $i$  prefers  $x$  to  $y$
- iv. Individual  $i$  prefers  $y$  to  $z$
  
- v. Individual  $i$  has the identity  $a$

In this example, it is important to note the order in which the statements are made. The implication is that an individual possesses certain features; in this case a specific preference ordering, that determines their identity. Individual  $i$  has the identity ' $a$ ' because they prefer  $x$  to  $y$  to  $z$ , they do not prefer  $x$  to  $y$  to  $z$  because they have the identity ' $a$ '. In this sense, identity is therefore depicted as a post-hoc concept, applicable for description and justification rather than for prediction. Interestingly, up until the last twenty years or so, this was in fact the predominant trend in empirical psychological research on the self, with, for instance, Wyhe's (1974) seminal review concluding that it was almost impossible to see the trend being reversed. Yet in the early 1980s, the treatment of the self-concept changed dramatically, with Suls (1982), Schlenker (1985) providing some early impetus behind what eventually became an apparent paradigmatic shift. Empirical studies began to reflect an idea that the self need no longer be envisaged as something that just 'reflected on-going behaviour, but instead mediated and regulated this behaviour' (Markus and Wurf 1987; p. 299). In other words, identity became a predictive variable, rather than an explanatory device, in studying human behaviour.

In terms of the example above, if identity is then denoted as the motivation behind human behaviour, the order of the statements then changes to the following:

- i. 'a' is a defined personal identity
- ii. Characteristics of a are preferring x to y, and y to z
- iii. Individual i has the identity a
- iv. Individual i prefers x to y
- v. Individual i prefers y to z

Crucially, the shift in empirical focus was also mirrored by a much more fundamental adjustment in the assumptions behind the stability of any given identity. Previously, the self was considered more akin to a generalised view of an individual – an average of all elements of that person (Block 1981; Costa and McCrae 1980), with some studies suggesting that individuals even take great pains to maintain and protect their respective self-conceptions (Swann and Hill 1982). But as with the argument above, the early 1980s again saw experimental psychology embrace a much more nuanced line of attack, with numerous new approaches to the issue all premised on the idea of multiplicity. Articulated using a plethora of different terms including images, schemas, prototypes, goals and even tasks, the core message remained that individuals are more than simply one self, they are in fact a combination of multiple selves.<sup>39</sup>

But how does this social psychological perspective on identity then reconcile with the rational choice informed approach that I developed in the previous chapter?<sup>40</sup> At this point, at least, two issues must be resolved: the first is that identity is now deemed a determinant of behaviour, and the second is that any given individual possesses many such identities simultaneously. Surprisingly though, neither statement proves impossible to deal with. Firstly, rational choice theory is very clearly able to encompass the notion that identity dictates behaviour. Indeed, the two schemes of propositions above make this point nicely, since the second contains no new information than the first. In this case, identity 'a' is the factor (or frame) which determines that individual i acts according to the preference ordering  $x > y > z$ . Classically, this is exactly what has

---

<sup>39</sup> Markus and Nurius (1986) also discuss the issue of *possible selves*.

<sup>40</sup> Chase (1992) also attempts to consider this issue, although he fails to really show how the models are actually compatible with each other.

been interpreted as the ‘rational economic man’ in the work of Becker (1976) or von Neumann and Morgenstern (1944). The second question, on how multiple identities might fit with decision theory, relies on an argument made a little earlier in the thesis. In chapter two, I showed how the notion of preference revision was entirely compatible with a methodology that relied on utility maximisation. Using Elster’s (1983) sour grapes for some examples of the many different ways in which revision might be possible, the argument on identity then becomes a logical extension of that conclusion. If each self has a designated set of preferences that are relevant to it, then for preferences to change one of two possibilities must be true. Either the preferences can change for a specified self, or the actual self itself can change. But imperatively, because identities are simply equivalent to preferences: these statements prove to be identical. In effect, if rational choice theory can allow preference revision, then it must also be able to cope with multiple personal identities.

### 3.2.2 DEFINING THE SELF: SOCIAL IDENTITY

Up to this point then, when defining the self, consideration has only really been made of the natural, or put another way, the personal, idiosyncratic properties of the individual in question. This reflects very strongly the principles espoused in the American tradition, and reconciles nicely with classical decision theory. The European tradition, on the other hand, reacting in part to the constraints of the former approach, has embraced a set of more collective features when defining the self (Farr 1981; Markova 1987). Born from this refocus, Tajfel (1974, 1978, with Turner 1986) offers an alternative picture of the self, one determined and represented by Social Identity Theory (SIT). Premised upon a distinction between types of situation that involve interpersonal, and those that involve group processes, the core motivation behind SIT is the idea that when participating in the latter, individuals are more likely to self identify in terms of the group of which they are a member. Put more simply, the self can be imagined to lie somewhere on a continuum, with wholly personal, and wholly social identity lying at each extreme:

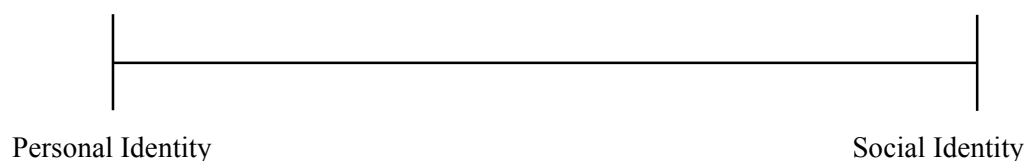


Figure 3.1: The Continuum of Identity

Irrespective of the way the relationship between these different types of identity can be modelled (see Hogg 1992), the fundamental assumption is that behaviour reflective of social identity displays a number of unique characteristic features. In-group bias, for example, is defined when individuals choose courses of action that are explicitly targeted to benefit a subset of individuals, of which they are a member, over another. Discrimination, where individuals actively make choices that punish individuals they consider outside of their group. Conformity<sup>41</sup>, where individuals ‘give in’ to social pressure and accept the ideas of the group over their own. Or stereotyping (Haslam et al 1999; Smith and Henry 1996), where individuals begin to take on the stereotypical properties/characteristics associated with the group. These, and a host of other characteristics are all typical of behaviour that is promoted by group identity. And quite crucially, as Bacharach (2006), Colman et al (2008) and Sugden (2003, 2008) all argue, as is agency transformation that induces individuals to act on behalf of the group – causing them to team reason.

If group membership can cause this very different type of identity and thus behaviour, a key issue clearly then becomes the study of such phenomena (see Brown 1988). Do all types of group have this effect of depersonalisation (Diener 1977; Oakes et al 1994), when individuals see themselves as being part of something ‘above’ their personalised self? It is obvious that throughout a person’s life they will come into contact with numerous different types of groups: from social groups, to sports teams, to political parties. All have very different characteristics, consisting of different members, social norms, rules and objectives. So how might they be classified in order to provide a meaningful addition to the social identity framework? Olson (1965), for instance, famously used the distinction of efficacy in order to identify small, latent and large groups - where the amount of impact that an individual could have on the group’s objective was the single criterion on which to distinguish between them. Hogg (1992), on the other hand, bases the distinction on the concept of cohesiveness (Festinger et al 1950), and asserts that it is possible to compare any group in terms of this scale. The more cohesive the group, the more likely it is to cause deindividuation and thus social

---

<sup>41</sup> Abrams et al (1990) distinguishes between compliance and conformity – the former implies simply bowing to the groups wishes in public, whilst the latter implies a much more true notion of acceptance of these ideas.

identity. Others, including Brewer and Harasty (1996), Hamilton and Sherman (1996) or Sherman et al (1999), take a similar approach to Hogg in terms of a variable scale, although discuss it with reference to the notion of entitativity. Conceptualising groups according to this latter concept is particularly relevant when the focus of the investigation is on team reasoning. In terms of social psychology, entitativity was first defined as the point where a collection of individual elements has the '*nature of an entity, of having real existence*' (Campbell 1958; p. 17). Whilst a number of different causal factors behind this process will be discussed a little further down, the point I seek to make right now relates to back to the consequences of such a process.

Team reasoning, which occurs when individuals ask themselves the question “what course of action should we take”, can clearly only happen once a particular group has been identified as a single entity – the ‘we’ must be clearly defined. Without the process of entitativity, neither Bacharach’s (2006) model of individuals reasoning on behalf of a team, nor Sugden’s (2000, 2003) argument of team preferences is possible. They both rely on the assumption that a collection of individuals can become a single entity beyond a mere aggregation of the members.

However, even with an appropriate scale of how to analyse groups in place, it does not yet provide a robust theoretical description as to how the processes of entitativity, deindividuation and social identity work. By merely stating that two types of identity (with characteristic behaviour) exist, and that being a member of a particular group is likely to trigger one or the other, a number of metatheoretical issues remain unaddressed. Firstly, it should be obvious that individuals will find themselves as members of many different groups throughout their lives. The concept of a single social identity, then, is far too limited – individuals clearly move between different social identities depending on the context. And secondly, because of this, any theory must then be able to offer some explanation or prediction as to *when* an individual is likely to identify in terms of one social identity over another. Finally, without these adjustments in place it is also clear that social identity theory offers no clarification on the notion of personal identity, thereby allowing any behaviour that cannot be explained by a particular social identity to be cast aside or ignored.

### 3.2.3 SOCIAL CATEGORISATION THEORY

Social categorisation theory (Turner 1982, 1985, et al 1987) attempts to build on social identity theory through addressing these issues. Fundamentally then, it proceeds on the following three claims:

- i. Individuals are able to categorise themselves at varying different levels of abstraction – from ‘humanity’ at one end, to ‘group’ in the middle, to the unique ‘person’ at the other.
- ii. The level at which individuals categorise themselves is determined by the level of salience of these varying different identities.
- iii. Salience of a particular self-category is determined by ‘accessibility x fit’ formulation (Oakes and Turner 1986).

These three assumptions agree with the proposition that a distinction needs to be made between personal and social identity. They rely on the concept of entitativity in so much as it is required for a group to be used as the unit of self-identification. But what it also offers, is a description of the cognitive process which underpins the discussion of why particular social identities are activated, or ‘in play’ at certain times.

The concept of salience is a particularly subtle one. Thomas Schelling (1960) famously used the term when considering the degree to which participants in coordination games managed to reach outcomes that classical rational choice theory did not support. By way of examples and metaphors, he used the term interchangeably with ‘prominence’, and suggested it as one reason as to why individuals converged on ‘focal points’ during such encounters. In the case where two individuals were tasked with meeting each other in New York without any agreement or discussion beforehand, the choice of Grand Central Station was chosen *because* it was deemed a more salient answer than, for instance, Times Square. But critically, the exact way in which the phenomenon of salience works, or indeed a strict definition, was never discussed explicitly.

Whilst experimental work in economics has been performed to test this idea more scientifically<sup>42</sup> (Mehta et al 1994), it is the conceptual literature within social psychology that again proves quite fruitful. Oakes and Turner (1986), for example, suggest salience can be split into two aspects. The first, accessibility, refers to the

---

<sup>42</sup> I use the term ‘scientific’ because Schelling famously critiqued his own work by labelling his experiments as ‘unscientific’.



‘readiness’ of an individual to accept a particular self-category, i.e. how much it fits with their previous goals, past experiences or current motives. Fit, on the other hand, has two further aspects (Oakes 1987): comparative and normative. The former involves a *metacontrast*, or put more simply, the degree to which the self-category in question is seen as being more cohesive than the alternatives on offer. The latter, normative, then extends this onto a scale. For example, to distinguish between two groups ‘a’ and ‘b’, individuals in ‘b’ must be both more similar to each other than in comparison with members of ‘a’, as well as be different in a very *specific way*. Factors that affect the salience of a particular social identity, then, can broadly be placed in one of these categories: they become determinants in the ‘accessibility x fit’ model.

Self-categorisation theory thus offers a robust conceptual model of identity formation, and describes the way in which individuals are able to move between multiple self-conceptualisations, depending on their relative levels of salience.

But before I proceed to discuss the various factors that have been found to act as determinants of salience, I wish to briefly discuss the extent to which this addition to the model of identity reconciles with the rational choice approach I drew upon in chapter two. The highly individualistic nature of a classical interpretation of decision theory seems quite antithetical to that of social identity. But, if there is nothing to stop individuals switching between utility functions, therefore preferences, therefore personal identity – surely there is nothing to prevent an individual acting on the basis of a utility function belonging to a particular group? Indeed, this seems to be nothing new. Margolis (1981) for instance argues that people cannot only switch between utility functions, but also between individual and social ones. Feeding this information back into the schema or reasoning format:

- i.  $a$  is a defined personal identity
- ii.  $b$  is a defined social identity
  
- iii. Characteristics of  $a$  are preferring  $x$  to  $y$ , and  $y$  to  $z$
- iv. Characteristics of  $b$  are preferring  $y$  to  $x$ , and  $z$  to  $y$
- v. Individual  $i$  has the identity  $b$
- vi. Individual  $i$  prefers  $x$  to  $y$
- vii. Individual  $i$  prefers  $z$  to  $y$

From a rational choice perspective, there is again nothing in this model that proves impossible to deal with. Yes, it might cause a reduction in the predictive power of positive political theory, but this is the entire reason for the addition of the social psychological framework – to offer predictions about where and when particular identities will become salient. In Bacharach's (1993, 1996, 2006) model of team reasoning, this is exactly what is meant by the term 'framing'. The rational choice element, as a theory of decision-making, then helps to create an experimental approach that allows for this to be observed. In short, combining both approaches leads to a more comprehensive theoretical way of examining and modelling human decision-making under these circumstances.

### 3.2.4 EXPERIMENTAL FACTORS PROMOTING SOCIAL IDENTITY

Now that a theoretical framework of social identity has been mapped, the next point I want to consider concerns the specific factors (or determinants in the accessibility x fit model) that have been found to trigger this self-categorisation in terms of a group. A considerable amount of experimental work has been done of this area (see Brewer and Miller 1996), with much in particular drawing upon social dilemma games (see Chase 1992) and other associated encounters. Very briefly, I want to consider a number of the most prominent factors that have been identified in this literature. These include belonging to an ad hoc group (Billig and Tajfel 1973; Locksley et al 1980; Tajfel 1970), belonging to the same social group (Dion 1973; De Cremer and van Vugt 1999; Kramer and Brewer 1984), the presence of an out-group (Kramer and Brewer 1984; Mullen et al 1992; Tajfel 1970), having common preferences, the use of common language (Dieckhoff 2004; Perdue et al 1990), having shared experiences (Drury et al 2009; Prentice and Miller 1992), face to face contact or discussion (Bornstein 1992; Dawes et al 1988, Orbell et al 1988; Wilder and Thompson 1980) and 'interdependence' (1985 Cookson 2000; Bacharach 2006; Sherif et al 1961; Turner 1981, 1982). Let me consider them each in a bit more depth.

#### *Belonging to an ad hoc group*

Some of the most early experiments on social identity were completed using the so-called 'minimal group paradigm', where individuals were assigned as members of arbitrary groups, by researchers under experimental conditions. Locksley et al (1980), for example, went as far as to make sure the individuals knew such an allocation was

entirely random, showing them lottery tickets to elucidate how the selection process worked. Even under these minimal conditions, participants continually displayed ethnocentric behaviour – discriminating in their choices to favour members of the same group they were assigned to at the beginning of the experiment. The conclusion from this evidence was quite clear then, social identity could be triggered by even the most trivial of group characteristics.

#### *Belonging to the same social group*

Following on from the minimal group concept, belonging to the same social group as other individuals has also been found to promote a social identity, and cause behaviour to favour the in-group. De Cremer and Van Vugt (1999), for example, used the distinction of whether participants were led to believe they were playing social dilemma games with either (i) generic university students, or (ii) those students at the same institution as themselves. Unsurprisingly, the results matched that of the minimal group studies, as players consistently favoured students from the same institution. A social identity, based on belonging to a particular social group, was clearly in evidence.

#### *The presence of an out-group*

Strongly related to the idea of an in-group is, of course, the reciprocal concept. Whilst it is quite possible for individuals to be members of a universal group (for example all *people* are part of the ‘human race’), membership is also often accompanied with the idea of an out-group: a collection of individuals who have formed an alternative, even competing association. In Tajfel’s (1970) original experiments then, this formed part of the actual setup, as behaviour was not only found to favour the in-group, but also crucially to discriminate against members of the out-group. By recognising a group that a particular individual does *not* self-identify with then, the features of the one they do becomes more salient, and therefore social identity in reference to that group is more readily primed.

#### *Common preferences*

One particularly strong determinant of group identification is, quite clearly, the presence of similar preferences or interests amongst a group. More often than not this is almost definitional, as many groups themselves are ordered, or convened on this basis. Personal preferences, on their own, are components of personal identity. But when individuals with the same personal identity are brought together in one group, and are

made aware of this fact, significant in-group bias is found to occur in their behaviour. Individuals doing so are clearly therefore treating the collective group as a single entity, and acting on a social identity primed by this common set of preferences.

#### *Use of common language*

Language has also been found to have a strong effect on identity – but on two slightly different levels. The first is on a more macro scale, where the use of a shared vocabulary or dialect creates a very meaningful and visible shared identity amongst a population (Dieckhoff 2004). The second concerns the actual content of the language. For example, Perdue et al (1990) tested whether the use of words explicitly connected with social identity, i.e. ‘us’, and ‘them’ affected how individuals ranked neutral ‘nonsense syllables’ to which they were coupled. The results were quite stark, as individuals consistently ranked those corresponding to an in-group more favourably than to those of an out-group – even though none showed any awareness of this ethnocentric behaviour/pattern.

#### *Having a shared experience*

Another factor that has been demonstrated to induce individuals to self-categorise in terms of a group has been the presence of a shared past experience. Drury et al (2009), for example, recruited individuals who had been involved in significant public disaster events, including the 1989 Hillsborough football stadium crush, and the Fatboy Slim beach party crush in 2002. By performing both a descriptive and qualitative analysis post interview, they conclude that there is significant support for the idea of a shared social identity amongst individuals from the same disaster, as well as potentially between individuals from comparative events. Past experiences, it therefore suggests, can cause deindividuation as the group, affected by the specific event (or type of event), becomes seen as a single social identity.

#### *Face to face contact or discussion*

The penultimate factor I wish to discuss is perhaps one of the most obvious, as it deals with a very visible characteristic of groups. Contact, either verbal or face to face, is one of the single most powerful explanatory variables for social identity - for instance, Festinger et al (1950) found that even proximity seems to produce some form of social bond between individuals. Experimentally, Wilder and Thompson (1980) looked at the behaviour of two groups of students assigned on the basis of college identity. They

found that discussion within the group led to heightened levels of in-group bias, as individuals began to identify with their relevant group more strongly. Conversely, contact between groups, where individuals were forced to interact with members of the other group, led to diminished social identity and less ethnocentric behaviour. Dawes et al (1988), asking individuals to play simple binary choice social dilemma games, also found significant evidence for the positive impact that discussion, in particular, has on social identity.

### *Interdependence*

The final factor to consider is perhaps the most complex, as it deals with the objectives of the group itself. Interdependence and the phrase ‘common interests’ are often used as synonyms, particularly in the social psychological literature. But when social dilemma games are used to investigate choice behaviour (particularly in experimental economics, see Cookson 2000), it becomes clear that a subtle distinction is required. To say that a group of individuals has common interests, is to say there are certain outcomes that might be beneficial to all those involved. But it says nothing as to how these goals can be achieved. Bacharach (2006; pp. 81-85) provides a particularly nice discussion of this debate, and suggests that whilst the definition of interdependence as ‘having common interests... that can only be achieved together’ (p. 84) might seem useful, it omits from consideration certain types of encounter. Mixed motive games, like the classic prisoners’ dilemma encounter, allow for the possibility that there are outcomes that might be best for both players when taken as a single entity, whilst simultaneously subordinate to another outcome when viewed under rational self interest. Because these games do not strictly conform to the general definition of common interests, interdependence is thus defined as a trait where a favoured outcome can only be achieved through the actions of all individuals, in cases where this behaviour is not assured. Experimentally, this effect has been consistently found to occur in public goods games before any treatment phase is initiated: individuals have played strategies associated with social identities simply on the basis of the game and payoffs offered (Andreoni 1988, 1990, 1995; Sugden 1984).<sup>43</sup>

Let me summarise where I have got up to in this chapter. By outlining a framework that provides a justification for the construction of social identity, which in turn links directly to the idea of ‘framing’ in Bacharach and Sugden’s models of team reasoning, I

---

<sup>43</sup> See Ledyard (1995) for a good survey of some of the most important early results in this field.

have completed the explanation of the rational choice side of the agency revision process. Or in other words, I have provided a generalised discussion of both *how* and *why* team reasoning occurs. The next step is to take a closer look at the extent to which deliberation might play a role in making social identity more salient for the individuals involved.

### **3.3 DELIBERATION AND SOCIAL IDENTITY**

---

I now return to the dual concepts of deliberation and deliberative revision. The claim that deliberation is likely to make individuals revise their agency and reason on behalf of a team is made on the basis that it will trigger individuals to self-define in terms of a specific social identity (Warren 1992). This leaves two related questions that must be answered in this final section of the third chapter. Do the features of deliberation discussed in chapter one match up to the arguments surrounding self-categorisation and social identity? And therefore, can the hypothesis made in chapter two, that deliberation causes people to team reason, be sustained on a comprehensive theoretical basis?

Before I do this though, I need to undertake some further path clearing and say something about what ‘type’ of deliberation I am talking about. In chapter one, I established a behavioural definition that was predicated on the idea of scale. On one end sat ‘ideal’ deliberation characterised by discussion that completely satisfied a number of normative criteria. At the other sat discussion that violated all the said maxims, and could not be termed deliberation in any meaningful sense. Between these points, however, is the idea of ‘better’ or ‘worse’ deliberation – dependent on the relaxation of certain criteria. In this thesis, my concern is to map the social psychological approach onto the *idealised* account of deliberation. Once I have done that, and following each point that I make, I will make a few suggestions as to how various relaxations might interact with these processes.

For the purposes of my argument, and to facilitate later analysis, this needs to be broken down into two different sub-sections. The first is to look at the key principles of deliberative behaviour, and to evaluate the extent to which these might cause an individual to self-identify in terms of the deliberating group. The second is to then draw

on the structural account, and see if this is more or less likely at any given point in the deliberation.

### 3.3.1 THE BEHAVIOURAL ACCOUNT OF DELIBERATION AND SOCIAL IDENTITY

As I outlined in chapter one, in the vast majority of the literature in political theory (and communication studies), deliberation is understood as a dynamic, communicative process between more than one individual. Without restating this debate, Mansbridge (2010; p. 65) defines it nicely as ‘communication that induces reflection on preferences, values and interests in a non-coercive fashion’. The first key word in this definition is communication. Deliberation is a very particular form of discussion, but it remains just that – discussion – and because of this, it has a number of features that reconcile nicely with social identity and self-categorisation theories. First, and most obviously, it can be compared to the results seen in the experiments of Wilder and Thompson (1980), Dawes et al (1988) and Dieckhoff (2004). Deliberation, as a discursive process that takes place within the group and utilises the same language and dialect, clearly has the potential to make a social identity more salient, as individuals begin to see the group they are communicating with as a single entity. Moreover, this is likely to be heightened by the classification of the individuals as being such a group by the process itself, whether participants really ‘buy into it’ on social group grounds (De Cremer and Van Vugt 1999), or on the basis of a simple ad hoc classification (Locksley et al 1980). Deliberation that satisfies the maxim of interactive communication then, reconciles nicely with the possibility of priming social identity. But what might happen when the maxim is relaxed ever so slightly? For example, take the case of online deliberation, where discussion is interactive but takes place behind the veil of a computer terminal. In this case, whilst a number of the same social identity comparisons can be made, it is also the case that the face-to-face element is lost. It is plausible to suggest, at least theoretically, that deliberation of this type might therefore have a lower propensity to trigger team reasoning.<sup>44</sup>

Moving now to the second comparison between deliberative behaviour and social identity. If the discussion reflects equality, mutual respect and reciprocity, then it is possible to argue that the link to social identity is strengthened further. Individuals that

---

<sup>44</sup> The degree to which any of these hypotheses regarding relaxations are true, is of course, an empirical question. Whilst it is not the focus of my project, which attempts to examine the effects of ‘ideal deliberation’, it is something I will discuss further in the final chapter regarding further research questions.

allow all other members to participate freely, and in doing so listen to and attempt to fully understand their points, are much more likely to find some degree of agreement with each other. Any form of agreement or common ground, then, is directly related to social identity. It allows individuals to describe the decision in terms of language that reflects entification of the group i.e. ‘we all agree on x, y and z’, in which case they share common preferences. Or even ‘we all agree on the scale to rank x, y and z’, in which case they share a common way of conceptualising these preferences. Either way, doing so can lead to deindividuation, and results in self-categorisation in terms of the group agreement and thus group identity. However, just as with point one, relaxing these maxims can have a significant impact on the potential for this effect. Consider what happens if mutual respect and reciprocity are relaxed. In this case, individuals are no longer required to empathise with other members of the deliberation. Removing this requirement reduces the possibility that any shared experiences, common preferences or common language might come to light. This, in turn, reduces the prospect that social identity will become more salient for the individuals involved. Again then, relaxing the idealised criterion has the potential to weaken the link between deliberation and social identity.

Thirdly, when deliberation is (i) based on reason (ii) made on the public principle and (iii) directed towards a final goal of decision-making, there are a number of ways it can be linked to social identity. The first, quite clearly, is through the use of language. Reason based argument involves the use of logical statements that are intended to prove a link between action and consequence, i.e. ‘choose x, so that y will occur’. In the arena of deliberation, the prioritisation of reasoned argument means that individuals are required to present claims in this format. Further still, when the public principle, incorporating Rawls’ idea of public reason, is introduced, it then becomes necessary to both present the argument linguistically in terms of the group, i.e. ‘we should choose policy x, so that outcome y will occur’, as well as secondly, to propose the policy that is most likely to benefit all the individuals concerned. In this case, the deliberative claim must be made in the format of ‘we should choose x, so that y will occur, because it is best for all of us’. These two aspects of reasoning can be traced directly back to two of the experimental factors discussed above. The use of the pronoun ‘we’ discussed in Perdue et al (1990), and the impact that shared interests and references of a group have. Both are present in deliberative reasoning of this nature, and thus it is logical to predict that it has the potential to make social identity more salient within the individuals.



However as with the previous two points, I also want to consider how a relaxation of these ideal deliberative criteria might impact on this debate. For example, consider deliberation over a topic that is characterised by deep disagreement, where pluralism dictates that complete satisfaction of the public principle is impossible. In this circumstance, the second best alternative is for individuals to rely on reasons that are characterised as ‘public spirited’ (Chambers 2004). Where two individuals already share common ground, this might indeed serve to prime social identity in the same way as a strict interpretation of public reason suggests. But in cases where two individuals occupy differing perspectives, both emanating from reasonable and popular positions, then it is possible to see how public-spirited reasoning might in fact work in the opposite direction. It is difficult to imagine, for example, a deliberation making social identity more salient amongst a single group consisting of individuals from different sides of the abortion debate. In short, deliberation in cases where public reason is impossible might do more to make personal identity more salient.

Finally, deliberation understood as a dynamic process between two individuals that incorporates these features, is clearly of a reciprocal nature. It is, for instance, impossible to have a deliberative conversation with a second individual if this person pays no attention to the norms of: (i) equality, because they talk too much; (ii) mutual respect, because they ignore what you say; (iii) reason centred argument, because they offer no justification for their claims; (iv) the public principle, because they talk only about what is good for their own self interest; or (v) decision-focussed approach, because they have no interest in ever reaching an agreement. I suggest then, that the single largest way in which deliberative behaviour is likely to make social identity more salient is through the concept of interdependency. For deliberation to occur between individuals, it is crucial for all those involved to behave in a very specific way. Just as with the prisoners’ dilemma game, where a socially optimal outcome exists that is only obtainable on the basis of behaviour that is not assured, high quality deliberation can only exist when all the participants adhere to these behavioural norms. When deliberation does occur, then, it becomes an almost self-fulfilling phenomenon. Individuals ‘x and y’ decide to deliberate according to these principles, and in doing so, become aware that it is only possible because they have both have participated in such a manner. This makes social identity more salient, which in turn motivates the individuals to maintain these behavioural norms.

Let me briefly summarise the main points I have made in this section. By examining the various behavioural criteria that idealised deliberation satisfies, it is possible to see how the agency revision argument put forward in chapter two might be substantiated on a theoretical basis. When discussion reflects all six of the criteria, the potential for social identity to become more salient than that of personal identity rises, and in turn, so does the propensity of individuals to employ team-reasoning. The next task is to look at how this effect might change during the different stages (or subtasks) of the deliberative process.

### 3.3.2 THE STRUCTURAL ACCOUNT OF DELIBERATION AND SOCIAL IDENTITY

#### *Education and Information Phase*

Moving on to the deliberative process, the first stage of education and information is characterised by the discovery and consideration of knowledge: where participants are expected to spend time learning about the issue(s). It is possible to think of education as a highly individualistic, personal endeavour – it is difficult, for instance, for more than one person to share the same copy of a text, and whether you understand a particular point is often dependent on your personal history, background or intellectual abilities. Conceived in this manner, it might seem that the first phase actually serves to promote personal, rather than social identity. But in a deliberation, especially on complex topics where the participants are selected *because* they have no prior interaction with the subject matter, the educational process is intended as a much more social, collective venture. Participants are often put in a situation where all members of the group are starting from the same point in terms of the issues under consideration, and because of this, the process to understand and clarify information becomes a collective activity. The focus, to put it bluntly, is on the group itself learning, rather than individuals gaining as much knowledge as they can.

Moreover, as I discussed in chapter one regarding the relaxation of the reason-based discussion criterion, the way information is presented in a deliberation also plays a substantive role. Whilst expert witnesses, briefing documents and other externally arranged information sessions are often conducted along highly rationalist lines (see Habermas 1984), the same is not true for personal testimony or whilst bearing witness (Barber 1984, Sanders 1997). The latter is a much more social form of communication, where emotion and empathy become relevant in the educative process. If a deliberation

includes this type of information, it allows for past experiences to come into play, and thus the possibility that a group identity might form on this basis. For example, if person 'a' relates a specific policy back to an event in their past that is shared by others, as Prentice and Miller (1992) or Drury et al (2009) demonstrate, individuals may begin to self-identify as belonging to this group. This is particularly relevant when a deliberation occurs over a topic that has had significant impact upon a number of the individuals involved – where members are able to recall past incidents that demonstrate the effect of these issues on their lives. Doing so constructs this shared experience, and therefore this phase has the potential to raise the salience of a social identity amongst the members. Because of this, whilst it might seem on first glance that the first task of the deliberative process is likely to work against raising the salience of a social identity, it simply isn't the case once a more thorough analysis is made.

#### *Identification of Solutions Phase*

The identification of solutions phase requires individuals to discuss as many different possible courses of action that might be taken to resolve the issue. Some options are already present in the discussion, because they come up either in the education/information phase, or because they are commonly known by the participants. Others are formulated as a direct result of the deliberation itself, where innovative ideas and solutions can be generated. In this case, there is a strong reason to think that deliberation may make social identity more salient. If a discursive group comes up with a proposal itself, and moreover that this proposal is identified as original, it is quite plausible for individuals to think of the group as a single entity. By doing so, and seeing themselves as part 'owner' of the policy, they have essentially become de-individualized in favour of the group. Consider a slightly more obvious example then. Imagine a collection of individuals in a park, when someone happens to produce a football. Players are randomly allocated onto different sides, and an impromptu game commences. Whilst the very act of classification serves to prime a social identity (individuals describe themselves as being members of team 'a' or team 'b'), this is further strengthened, for example, when one team scores. It is plausible for participants of the scoring team to describe the incident in one of two ways: either 'individual x scored a goal' or 'our team scored a goal'. With the latter, which is often the case when a goal involves many members to score, it is clear that the individual is responding from the relevant social identity. The exact same process is at work when deliberation leads to a new proposal made by the participants. It is possible, and even likely, for

individuals to describe the proposal as ‘ours’ referring to the deliberative group – at which point they are giving the group entativity, and therefore raising the salience of a particular social identity. Again then, the identification phase also seems a plausible element of the deliberative process to help trigger individuals to self-conceptualise in terms of a social identity.

### *Evaluative Criteria Phase*

Once various solutions have been identified and understood by all the participants, the next stage in the deliberative process is for individuals to establish a way of evaluating the different proposals. Because of the way deliberation requires individuals to make claims regarding their preferred policies (the public principle), it is easy to see that the very task is one of cooperation and public reason. Individuals must work together in order to find an acceptable set of criteria upon which to analyse the policy proposals – drawing upon both the previous sessions. If individuals are expected to work together as a group to come up with their own ranking, then it is clear that a number of factors will be at work. Interpersonal communication (Wilder and Thompson 1980) will be used in order to come to an agreement, and shared experiences (Prentice and Miller 1992) will be drawn upon in order to elicit cooperation. But just as important as all of these, it is also the first time the group must come to an explicit *decision* over a set of possible outcomes.

Following the same argument suggested in the section on equality and mutual respect above, it is crucial to acknowledge that the requirement of this phase is *not* to secure agreement on the policy decision/recommendation. The evaluative criteria element of a deliberation only requires individuals to make a decision on the *way* in which the various proposals identified in the previous phase are ranked. Metaconsensus, as it is termed in Niemeyer and Dryzek’s paper (2007), is therefore a much weaker level of agreement for the group to reach. Whilst it is correct to suggest that the act of making a decision as a group has the potential to make the group/social identity more salient, the level at which this happens in this phase also means the impact should be much less pronounced than when compared to decision-making over actual policy proposals. For a group member to say ‘we choose policy proposal a’, it is clear that the group must have a level of entativity, as well as the possibility that the identification of a common preference position has occurred - found to trigger social identity in experimental work. Yet it is much more difficult to imagine the statement ‘we all agree on the process of

how the policy proposals should be ranked’ as achieving the same outcome. Indeed, the extent to which metaconsensus might affect individual behaviour in group situations is an emerging area: where there has been no experimental work designed to explicitly test this hypothesis up to this point. Conceptually, at the very least, the point does appear justifiable.

To briefly restate then, the evaluative criteria phase certainly seems to have a number of features within it that are likely to raise the salience of group/social identity. What is apparent more in this section than any of the previous two, however, is that the *degree* to which this is likely is very difficult to predict. I’ll go into this in more depth in section 3.3 below.

#### *Decision-making Phase*

Although the more controversial of the four stages of the deliberative process, I want to look at the decision-making phase for two reasons. The first concerns a simple restatement of the way in which some form of agreement is likely to trigger entitativity and therefore social identity – covered in the equality/reciprocity section on deliberative behaviour, and the evaluative criteria section of deliberative phases. The second way, however, concerns the point at which the decision is made. In chapter one I discussed the way in which the concept of consensus had been used in deliberative democratic literature, and specifically, that a debate existed as to what the ends of an ideal deliberation should be. Some argued explicitly for unanimous agreement, whilst others suggested this as oppressive or simply impossible. However, the conclusion that I arrived at, based on a distinction between ‘objective’ and ‘result’, is crucial here. By setting a consensus as the objective of the decision-making phase, whilst simultaneously acknowledging the ‘fall back’ of a majoritarian outcome, this element is more likely to involve discussion that reflects points (iii) to (vi) in the key principles of deliberative behaviour. Because of this, it is therefore more likely to have a similar impact on social identity that these features predict.

However, the very fact that a decision must be made at this point also introduces a concern that less deliberative behaviour is perhaps made more likely. By this, I mean the extent to which individuals might relax the ideal criteria in their attempts to make the final decision more closely match their preferred outcome. In effect, it is possible to imagine a situation where the decision-making phase prompts those involved in the

deliberation to act strategically. Again then, consider a topic categorised by deep disagreement such as the abortion debate referred to earlier. If a group of individuals reflective of all sides of the issue are required to agree on a single set of policy proposals (i.e. legalisation or criminalisation), then a consensus is impossible. This is very different from the situation where consensus is possible but improbable, as it now removes the incentive for individuals to make deliberative claims according to the public principle. Why bother presenting an argument in terms of ‘x is best for all of us’ if you know half the group believe ‘y is best for all of us’ and will never change their opinion. In this type of deliberation, and particularly at this stage of the process, it is therefore possible to imagine personal rather than social identity of the group becoming more salient for those involved. I will return to this specific point in chapter six of the thesis.

### **3.4 CONCLUDING REMARKS**

---

The objective of this chapter has been to provide a more robust theoretical foundation for the claim made in chapter two, that deliberation might trigger agency revision, and cause individuals to team reason. By engaging with the social psychological literature on social identity, I have provided an argument that both reconciles with rational choice theory (through Bacharach’s notion of framing) as well as reconciling with key principles of deliberation and deliberative behaviour.

With the theoretical model now firmly in place, the next step is to examine the extent to which practice might live up to this theory. In the next two chapters of the thesis I now turn to the question of how this phenomenon might be investigated empirically.

# — CHAPTER 4 —

## MEASURING TEAM REASONING: AN EXPERIMENTAL ECONOMICS METHODOLOGY

### 4.1 INTRODUCTION

---

In chapter two I presented a three-fold analytical model to which the concept of deliberative revision can be applied to decision making, with the rational choice approach to political science used as a tool to elucidate a number of key supporting arguments offered by deliberative theorists. Two of these levels, namely issue and preference, have been the subject(s) of the recent empirical turn in deliberative democratic theory. The third, agency, has largely been ignored. Having demonstrated the applicability of such a revision to deliberative claims in chapters two and three, I now progress to the experimental section of this thesis. To what extent does the hypothesis regarding deliberation and agency revision hold true empirically?

This chapter therefore has the following two objectives:

- i. To develop an appropriate method through which the phenomenon of agency revision, leading to team reasoning, can be investigated in relation to political deliberation.
- ii. To provide the details of the *experimental economics* side of the investigation, in particular the structure of the games and the specific instructions that were given to participants.

Once both these objectives have been completed, in chapter five I will then discuss the details of the deliberation that was used as a case study.

### 4.2 WHY EXPERIMENTAL ECONOMICS?

---

Experimental economics, the empirical, investigative arm of rational choice theory, usually considers its subject matter in one of two ways. Originally, its role was constrained merely to the testing of pre-existing theories of rationality. It was,

essentially, an endeavour to discover whether actual human decision-making conformed to the rigid assumptions of microeconomic theory. Individuals were placed in experiments designed to approximate economic decision contexts, given choices to make, and the results were then compared directly to theoretical prediction. Yet once consistent discrepancies were found between classical economic prediction and reality, for example the work referred to earlier of Kahneman and Tversky (1979) or Loomes and Sugden (1982), its role shifted dramatically. Experimental economics then established, and now currently fuels, a large and expanding literature on what Starmer (2000) terms ‘the hunt for a [universal] descriptive theory of choice under risk’.

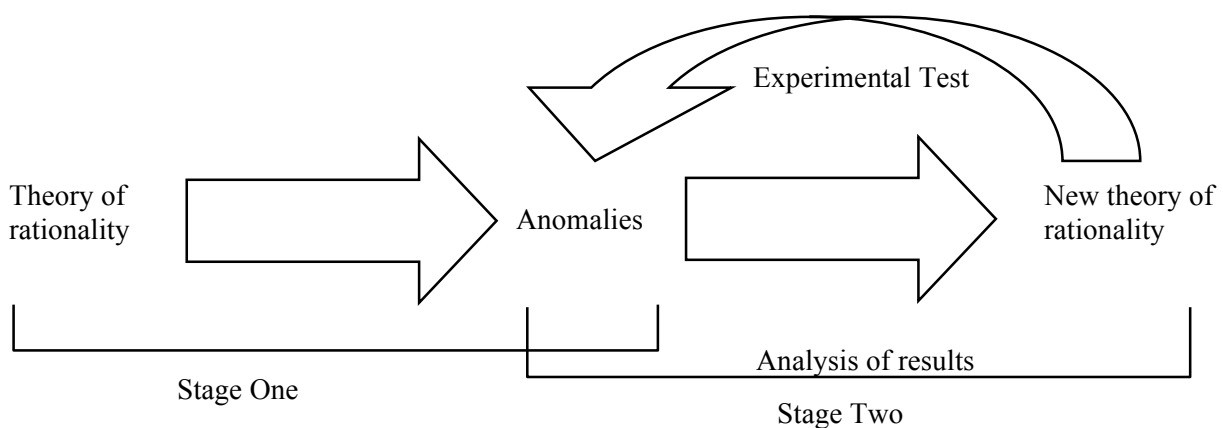


Figure 4.1: The Methodology of Experiment Economics

As a discipline in its own right, the research agenda within experimental economics has predominantly remained fixed on the second stage of the diagram – each new ‘universal’ theory of rational choice ultimately proving not so ‘universal’ as robust anomalies are identified. However, the investigative portion of this thesis does not, it is important to state, share the same stated end goal as experimental economics. The argument that deliberative mechanisms, such as citizens’ juries, transform agency and thus induce individuals to team-reason does not mean that either (i) it is the only transformation of how people reason that might take place, or (ii) that if it does not, the entire notion of team reasoning should be abandoned as concept entirely. What is important for this study is simply *whether* deliberative mechanisms induce this specific shift.



There are, then, a number of significant reasons as to why the experimental economic approach to social science proves most suitable to the task of investigating agency revision due to deliberation. These can be apportioned into five separate arguments:

- i. Observation ability
- ii. Clarity
- iii. Spill-over effects
- iv. Issue effects
- v. Agency isolation

Dealing with the issue of observation ability first. In reality, preferences and agency cannot be examined directly. There is no device that can simply ‘scan’ an individual to measure what they like or dislike, and more so, how they act upon these desires. Further still, merely asking individuals to surrender such data, for example in simple question form, is also highly problematic as well as unreliable. Experimental economic methodology, though, works by utilising the link between choices, and the preferences and modes of reasoning (agency) that underpin them. These choices can be observed, and therefore preferences and modes of reasoning are uncovered in an indirect manner. As Colman et al (2008) put it:

Neither preferences nor modes of reasoning can be observed directly, but predictions can be made about choices that would result from collective utility maximisation and team reasoning, and that behaviour can be observed directly.

(Colman et al, 2008; p. 4)

Secondly (and related significantly to the first point), is that it should be apparent that experimental economics is not the only approach that works on the principle that claims choices can reveal preferences and agency. Q methodology, for example, takes statements that are reflective of different policy positions and asks individuals to rank them in order of agreement. Preference orderings, and to an extent, how people reason on them, can then be inferred. Yet approaches to investigating the effects of deliberation that are based on real-life examples, with complex policy alternatives, provide fundamental problems with discovering the true nature of preference and agency revision. Interpersonal comparisons of utility that are necessary for any concept of pro-sociality to work become increasingly difficult to make when dealing with

outcomes that affect individuals in such different ways. For example, consider an individual (post deliberation) with the following new utility function over a decision  $x$ :

$$u_i = \alpha[u_i(x)] + (1 - \alpha)[u_j(x)]$$

The utility that individual  $i$  gains from the choice of  $x$  is, as I have discussed in chapter two, now dependent on the utility that individual  $j$  in turn receives. There are then, two clear difficulties. For the individuals involved there is an information gap to bridge: what value does individual  $i$  denote to the term  $[u_j(x)]$  in their utility function? In multifaceted policy decision-making, where in this case the choice of  $x$  may have unknown (to  $i$ ) consequences for  $j$ , the estimation of such a value becomes increasingly difficult. Secondly, this difficulty is multiplied even more for an empirical political scientist trying to measure the extent to which, for instance, preferences do become more ‘other-regarding’. Without a clear transposition between outcome and utility, making interpersonal comparisons in some sense ‘meaningful’, no such technical conclusions can be made. By using monetary values to represent outcomes then, a larger degree of clarity is secured: it is, for example, reasonable to assume the individuals prefer more money to less, and since money is universal (it can be spent on anything the individual desires), it is also an extremely useful proxy for utility. Interpersonal comparisons thus become much easier for the individuals within the experiment to make, and importantly for the political scientist, to observe and make meaningful conclusions from.

The third argument for using an experimental economic approach, then, concerns the status of the issue in deliberative revision. Many of the benefits that deliberation is argued to yield have no real link to the problem being considered. Individuals who partake in citizens’ juries, for instance, are expected to leave them not only with revised preferences and agency on the topic at hand, but as different, or ‘better citizens’ (Fishkin 1995). If this is the case, then these ‘spill over effects’ – whereby deliberation over one issue causes individuals to approach others from the same perspective – are undeniably an important area for empirical investigation. The nature and set up of experimental encounters within economics, in particular the use of artificially constructed games, thus provides a clear separation between the topic that is discussed in a deliberation, and the decisions which individuals are then required to make. This, it is important to note, is especially relevant to the debate that surrounds the impossibility

of institutionalising deliberative mechanisms within a large democratic society. If some of the benefits that mechanisms like citizens' juries secure can be achieved by only attending one such forum, then the implications for the macro-arguments are significant. This is a matter I will return to in chapter seven.

Fourthly, and perhaps the most important argument I have made so far, concerns the influence of the issue itself on the nature of a deliberative revision. As I demonstrated in chapter two, there are some convincing arguments for why a revision of the issue being discussed in a deliberative mechanism can benefit the decision-makers in question. Generating entirely new alternative policies that may not have existed before hand, or allowing compromise positions not originally designated options are but two examples. In such cases, it therefore becomes extremely difficult to truly capture how preferences and agency have changed: the new preferences and modes of reasoning may simply reflect what was true (but not available) in the first place. By keeping the issue constant, so that individuals are faced with exactly the same issue and associated decisions to make both before and after they partake in a deliberation, these inconsistencies are guarded against. A true, or in some sense more 'pure' result of the deliberation can be identified.

Finally then, just as this fourth point demonstrated the ability of an experimental economic methodology to remove the effects of issue revision from the investigative process (to truly capture preference and agency revision), the fifth point is concerned with how the same logic can be applied in order to distinguish between these two remaining levels. Yet largely as a result of the highly subtle conceptual distinction between individuals who "I" reason with pro-social preferences, and those who team-reason, this is a much more complex task.

#### 4.2.1 ISOLATING AGENCY REVISION FROM PREFERENCE REVISION

In discussing the benefits of agency revision in chapter two, I suggested that triggering individuals to employ team reasoning should be considered beneficial for society, since they will voluntarily participate in the provision of socially valuable public goods. It might seem, then, that all that is required to test the presence of team reasoning is to reverse this logic, and place individuals into exactly these types of games. If they choose the strategy characterised as defect in a one shot prisoners' dilemma encounter

before deliberation, and cooperate after; then there are grounds to argue that the process has caused agency revision in this manner.<sup>45</sup>

However, whilst it is likely that some individuals who choose the strategy cooperate in many prisoners' dilemma games are doing so as a result of having team reasoned, it is also quite plausible that some are not. Some prisoners' dilemma games can be 'solved' merely by the presence of other-regarding preferences (Basu 2006). In these cases, it is impossible to ascertain from the change of strategy choices whether agency revision has definitely occurred. The objective, then, is to find a game theoretic encounter with a collectively rational outcome that cannot be the product of a choice other than one arrived at by an individual who also team-reasons. One such encounter then, as I described in chapter two, is the Hi-Lo game featured heavily in both Sugden and Bacharach's work. Here, when individuals partake in the game on the basis of self-interested preferences, two Nash equilibria at [high, high] and [low, low] exist. Yet from a 'we' perspective (and the associated maxim of payoff dominance), [high, high] is the sole rational outcome:

		Individual <i>j</i>	
		high	low
Individual <i>i</i>	high	10 , 10	0 , 0
	low	0 , 0	1 , 1

Figure 2.3 (restated): The Hi-Lo Game

The question, of course, is why can't high be deemed a rational strategy in Hi-Lo by appealing to other-regarding preferences? Suppose, then, that individual *i* now shifts to a utility function with an altruistic concern, representative of rational altruism (Bardsley and Sugden 2006) or pro-social preferences (Van Lange 1999). More specifically, player *i* now considers player *j*'s payoffs to be equally important to their own, so that their utility function is depicted by:

$$u_i = 0.5[u_i(x)] + 0.5[u_j(x)]$$

<sup>45</sup> Of course, it is also worth noting that some individuals, when confronted with public goods games, often play the cooperate strategy as a simple matter of course – these individuals do not require a deliberative mechanism to induce such a shift (see Frank 1988 and Marwell and Ames 1981). I will consider this debate in relation to 'ceiling effects' below.

Due to the symmetric structure of the game however, the resultant other-regarding payoffs are merely the same as those from an entirely self-interested perspective. Looking at the [high, high] outcome, for example:

$$u_i = 0.5[10] + 0.5[10]$$

$$u_i = 10$$

In fact, given *any* combination of weighting between the altruistic and self-interested elements, precisely the same result obtains. Preference revision does not alter the structure of the interaction, and thus cannot rationally explain why the strategy high should be played consistently over that of low. Further still, an identical result is also obtained when a second common derivation of this pure ‘co-ordination’ game, the Stag Hunt, is played:

		Individual $j$	
		left	right
Individual $i$	left	10 , 10	0 , 8
	right	8 , 0	1 , 1

Figure 4.2: The Stag Hunt

From the deduction that individuals who choose high in Hi-Low and left in the Stag Hunt do so only as a result of employing team reasoning, then it might seem that both games would be useful in investigating the impact of deliberation. However, two significant problems arise with both: one practical, and one more technical. The first is that when played experimentally, almost all players automatically choose high and left respectively. At the same time as being used as a justification within game theory for the application of team reasoning to economic interactions then, both games are therefore also highly unsuitable for investigating agency revision in deliberation. The structure and payoffs of the game itself are enough to induce all individuals to approach the game from a “we” perspective: in terms of the three-fold model developed earlier, the *issue* triggers *agency* revision without the need for any deliberation. The second problem with using either the Hi-Lo or Stag Hunt encounter, is that whilst individual, self-interested reasoning does not offer an argument as to why high or left are played consistently, it also does not offer a definitive reason to why they should *not* be played. Since in each game two Nash equilibria occur, at least some strategy choices of high and

left can therefore be explained on the basis of self-interested preferences and individual agency. Any result that deliberative revision may have, then, becomes obscured. It is, essentially, impossible to definitively conclude why a strategy choice has been made.

Because of these foundational problems, i.e. the need to find a game that requires agency revision (from I to we) to induce a particular strategy to be played, and the strong framing effects of many games with collectively rational outcomes, the actual construction of the interactions becomes a quite technical endeavour. Colman et al (2008), in the first experimental paper explicitly investigating the phenomenon, offer an extremely useful starting point in their second section. In fact, as Sugden (2008) argues in a related commentary, their fifth interaction in particular provides a powerful comparison between how individuals who reason from a “I”, and those who reason from a “we” perspective play the game. In their investigation, individuals were presented with the following written instructions (Colman et al 2008; p. 7-8):

You are now going to make [several] decisions, from which you can earn more money. There are no scenarios with these – they are purely cash decisions. You and the other person will be presented with the identical problems. To work out the likely consequences of any decision, you will have to take into account what the other person is likely to choose. Once again, one of these problems will be chosen at random by a computer, and you and the other person will receive the amounts shown, in cash, depending on both your choices for that problem.

And asked to play the following game:

		Individual $j$		
		C	D	E
Individual $i$	C	8 , 8	5 , 5	5 , 9
	D	5 , 5	6 , 6	6 , 7
	E	9 , 5	7 , 6	7 , 7

Fig 4.3: A 3x3 Game Theoretic Interaction

From a classical (self-interested and individual) perspective, only one Nash equilibrium at [E, E] can be sustained: strategy E is the best reply for each individual to make given that the other is also playing E. Nash reasoning therefore argues that rational players will follow such logic. But not only that, it is also the case that E is a dominant strategy

for both players – meaning strategy E choices are the product of self-interested preferences without the more stringent requirement of Nash reasoning. By symmetry, then, the following is true for both individuals:

- i. If player  $i$  plays C, player  $j$  should play E since 9 is greater than 5 (D) and 8 (C).
- ii. If player  $i$  plays D, player  $j$  should play E since 7 is greater than 6 (D) and 5 (C).
- iii. If player  $i$  plays E, player  $j$  should play E since 7 is greater than 6 (D) and 5 (C).

Given this result, it would seem that any individual who plays strategy E before engaging in a deliberative process, and C thereafter (where [C, C] is evidently a collectively rational outcome), does so on the basis of having undergone agency revision to that of a “we” perspective. But to see whether this statement entirely holds true, it must be ascertained whether the choice of strategy C can be explained by anything other than an individual who employs such reasoning? Can deliberative revision that produces other-regarding preferences, for example, transform the payoffs of the game to sustain an equilibrium at [C, C], thus making strategy C choices rational? If this is impossible, then the game successfully distinguishes between preference and agency, and depending on the framing effects of its construction, would provide a useful experimental game for studying this specific impact of deliberation. Formally then, if other-regarding preferences can explain C choices, then the following statement must be true,

- i. Given a utility function in the form of  $u_i = \alpha[u_i(x)] + (1 - \alpha)[u_j(x)]$ , there is a value of  $\alpha$  that yields a Nash equilibrium at [C, C].

The following payoff functions can be constructed for player  $i$ , dependent on the choices of player  $j$ :

$$\begin{aligned}
 u(C,C) &= \alpha(8) + (1 - \alpha)8 = 8 \\
 u(E,C) &= \alpha(9) + (1 - \alpha)5 = 4\alpha + 5 \\
 u(D,C) &= \alpha(5) + (1 - \alpha)5 = 5
 \end{aligned}$$

Re-arranging a little shows that when the following is true, C becomes the rational strategy to play when agent  $j$  also plays C:

$$\alpha < 0.75$$

Which is exactly the same as saying that when the ‘other-regarding’ weighting in the utility function is greater than 0.25, then by symmetry, [C, C] is a Nash equilibrium. It would seem from this result, that the strategy which team reasoning equates to is also the strategy that a particular set of other-regarding preferences might rationally explain. Or in other words, that both agency and preference revision can explain the choice of C in this game. Rather crucially for this methodology, however, this assertion is incorrect. The above game may have the property such that when  $\alpha < 0.75$  then an individual maximising  $\alpha[u_i(x)] + (1 - \alpha)[u_j(x)]$  will do best by choosing strategy C. But in this game, an individual simply cannot maximise this function because they do not know what strategy the other player will choose; the game is what Sugden (2008) calls ‘non-decomposable’.

To understand specifically what this means, consider another simple prisoners’ dilemma, with two players  $i$  and  $j$ , and two strategies for each player. Player  $i$  must choose between up and down, and player  $j$  between left and right. The most common way to describe this type of game is in a simple 2x2 matrix format as follows:

		Individual $j$	
		left	right
Individual $i$	up	1 , 1	-1 , 2
	down	2 , -1	0 , 0

Figure 4.4: A Decomposable Prisoners’ Dilemma

This game, though, can be articulated in a different manner. If player  $i$  chooses up, no matter what  $j$  does, he loses 1 (either going from 2 to 1, or 0 to -1) and  $j$  will gain 2 (going from -1 to 1, or 0 to 2). Similarly, if player  $j$  chooses left, then she loses 1 (going from 2 to 1, or 0 to -1), and  $i$  gains 2 (going from -1 to 1, or 0 to 2). Put more succinctly still,  $i$  chooses between up =  $\{-1, +2\}$  and down =  $\{0,0\}$ . Player  $j$  then chooses between left =  $\{+2, -1\}$  and right =  $\{0,0\}$  where the first entry in each vector is the net change in payoff to  $i$ , and the second entry the net change in payoff to  $j$ .



i. Vector table for player  $i$ :

	Payoff Effect	
	Player $i$	Player $j$
up	-1	+2
down	0	0

ii. Vector table for player  $j$ :

	Payoff Effect	
	Player $i$	Player $j$
right	+2	-1
left	0	0

In a decomposable game such as this example, each individual can then therefore decide what strategy to choose simply by looking at his or her own ‘effect’ vectors. For example, with the following other-regarding preferences, where the utility function is given by  $0.5(u_i) + 0.5(u_j)$ , individual  $i$  can simply work out their best course of action by applying the function to their two vectors, and seeing which one yields the greatest increase in payoff:

$$[0.5u(i) + 0.5u(j)] \longrightarrow \{-1, +2\} = +0.5$$

$$[0.5u(i) + 0.5u(j)] \longrightarrow \{0, 0\} = 0$$

In this type of prisoners’ dilemma, other-regarding preferences can therefore be used to rationally explain choices that are also indicative of a team-reasoned strategy. From a methodological point of view, the imperative question to consider is therefore how to distinguish between games that can be decomposed (and are thus unsuitable for testing the presence of team-reasoning), and those like Colman et al (2008)’s 3x3 game above, that cannot. The answer, to put it most simply, lies in the symmetry within the payoffs in the game. For instance in figure 4.4, when individual  $j$  is playing ‘left’, individual  $i$  does better by playing ‘down’, with an increase of 1 in their payoff. Imperatively, the identical argument is also true when individual  $j$  plays ‘right’. It is this symmetry in payoff changes that allows the decision vector for player  $i$  to be entirely independent of

the strategy taken by player  $j$ , allowing the game to be decomposed. To prevent this, and to retain the strategic nature of the interaction, the differences in payoffs across the available strategies must vary. For example, consider the following alternative prisoners' dilemma:

		Individual $j$	
		left	right
Individual $i$	up	1 , 1	-1 , 2
	down	3 , -1	0 , 0

Figure 4.5: A Non Decomposable Prisoners' Dilemma

Now, the difference in payoff for player  $i$  when they play down is dependent on the strategy that player  $j$  takes: either  $\{+2\}$ , from 1 to 3, or  $\{+1\}$ , from -1 to 0. The interaction therefore *cannot* be decomposed into simple independent decision vectors for each player. For completeness, then, consider again Colman et al (2008)'s 3x3 game, it is quite obvious that the differences between payoffs for  $i$  are not consistent across the strategy choices of  $j$ . Indeed, the differences when  $j$  is playing C are  $\{-3, +4\}$ , when  $j$  is playing D are  $\{+1, +1\}$  and when  $j$  is playing E are  $\{+1, +1\}$ . The payoffs to  $i$  are thus totally dependent on the choice of  $j$ . Since  $i$  has no independent reason to believe  $j$  will play C, then they are in fact unable to maximise an other-regarding utility function, and thus pro-social preferences cannot be used as a rational justification for either player to choose C. It is, therefore, an example of a game that can be used to test whether an individual is team reasoning, and thus represents a case in point for the usefulness of an experimental economic methodology to this project. In short, non-decomposable games are suitable for investigating whether individuals are employing team reasoning, rather than just acting on other-regarding preferences.

### 4.3 PROBLEMS WITH THE EXPERIMENTAL ECONOMICS APPROACH

Having discussed a number of arguments in favour of an experimental economic methodology, I also want to answer some possible critiques than might be levelled at such an approach. For reasons of simplicity, they can be more or less split into two broad categories: those which fundamentally attack the experimental approach to social science in general, and those which are specific problems associated with combining the

methodology to the study of deliberation. Let me start with the more general problems, before moving on to those specific to this project.

i. Internal and external validity

Internal validity refers to the ability of the researcher to make accurate causal conclusions from empirical data; or to put it another way, the truthfulness of the proposed relationship between cause and effect within the study. Relating this purely to experimental economics, this concern can be interpreted in perhaps two different ways. Most fundamentally, it examines the link between preferences and modes of reasoning, and the choices that result from them – questioning the very basis of *what* makes an individual act. More practically though, it manifests itself in the extent to which the conditions of the experiment match the assumptions of the theory or model being tested – it is a matter of experimental design. In this sense, the most crucial question that internal validity poses is whether certain strategy choices that are played within game theoretic interactions can be attributed to the specific types of preferences and modes of reasoning that are being investigated. Essentially, this requires a simple restatement of the arguments posited directly above, where the isolation of agency transformation (from issue and preference) is made possible. The notion of internal validity, thus, is not only a rather vapid critique, but in fact turns out to be a particular strength of the approach.

Regarding external validity, the critique surrounds the extent to which data, and conclusions based upon it, can be generalised to a wider context (Lowenstein 1999). Are studies completed in a laboratory applicable to examples within society more generally? Most political scientists (or even economists for that matter) from a behavioural tradition would argue quite vociferously that it is here where the experimental economic approach falters. When individuals are asked to act in *true* political contexts, such as voting in a referendum, then the process by which they make their decision is markedly different from that when making decisions over monetary payoffs. The claim isn't that that experimental economics approach offers no useful data (for this would be an internal validity argument), but rather that the context is so artificial (and specific) that conclusions based upon evidence it provides simply cannot be applied to any other situation.

For experimental economics, quite clearly, the trade-off between internal and external validity is a critical area for consideration. What degree of reality must be sacrificed in order to restrict the number of variables impacting the process under investigation? It is this logic, or in some sense a cost-benefit calculation, which dictates the appropriateness of the methodology for this project. Because of the highly subtle nature of the distinction between outcomes that are the product of individuals employing individual reasoning with other-regarding preferences, and those that are the product of individuals employing team reasoning, then an approach which utilises monetary payoffs is the most precise way to grasp the technical structure of preferences and modes of reasoning that ‘are in play’. Moreover, the laboratory conditions of the experiment, by reducing the influence that other variables might have, should in fact offer a much more robust conclusion as to the impact of political deliberation specifically on agency revision.

It is important to make clear though, that removing variables is *not* the same as removing the context in which decisions are made.<sup>46</sup> In particular, a common argument levelled against the experimental economic approach, using the concept of external validity, is that decision-making over money itself often causes individuals to become unnaturally self-interested. But fundamentally, this investigation makes no claims regarding the natural level of pro-sociality of human beings. The initial preferences of an individual, and the mode of reasoning used from them, are in one sense irrelevant – it is the change between pre and post deliberation that is important. In fact, since one of the most significant problems with using these types of games is the strong framing effect of the collectively rational outcome, then ensuring the context can prevent this automatic agency revision proves pivotal. The apparent tension between internal and external validity within experimental economics as a discipline then, tends to suggest a sacrifice of the latter to ensure the former. Yet, because the research question of this thesis is directed towards an entirely different goal (investigating the impact of deliberation on agency, rather than finding a universal theory of choice under risk), the same conclusions do not obtain. Indeed, it appears that a strong case can be made for the approach offering a more, not less universal, conclusion as to the effect that deliberation has.

---

<sup>46</sup> It is important to distinguish between the decision context (i.e. the problem that an individual is required to make a decision over), and the more general notion of context, used here, to denote the general conditions (i.e. location, the *type* of decision being made) of the experiment.

With the more general debate regarding the use of an experimental methodology in social science considered, it is now necessary to confront some of the more specific problems in applying this approach to the study of political deliberation. I suggest these can be broken down into five different points, which I now want to consider in turn:

- i. Status quo bias (testing and instrumentation threat)
- ii. Individual-type bias (selection threat)
- iii. Situation-type bias (selection threat)
- iv. Incentives
- v. Understanding (game complexity)

Investigating the impact that deliberation has on how people reason, by its very nature, requires a ‘pre-test/post-test’ research design. Individuals must play games both before and after they participate in a deliberation. One of the features of the experimental economic approach, as I stated earlier, is that it can remove the effect that a revision of the issue being considered might have on individual behaviour – since the interactions (payoffs and strategies) are fixed and not subject to the forces of change. As a consequence though, this allows for the possibility that a status quo bias may arise.<sup>47</sup> Individuals, faced a second time with the same set of decisions to make, with the same information and same decision making context, are at least partially likely to simply follow their original choices. Path dependency, in this sense, could be triggered by a desire to look consistent in the eyes of the researcher, a desire to look consistent in the eyes of the other players, or even a misunderstanding that somehow, consistent behaviour will be rewarded. To guard against these three potential problems, experimental economics has established a number of conventions, including strict anonymity between the decision-maker and the researcher, as well as between the decision-maker and their assigned partner (Hoffman et al 1994, Lowenstein 1999). Moreover, written and verbal instructions are presented in a manner so as to make clear that each game is entirely independent from another, and that a strategy choice in one interaction has no impact on the strategy choice in another. Specifically, the information that is given out before and after the deliberative mechanisms are undertaken must stress all these points in order to ensure that decisions are made only on the basis of current preferences and current modes of reasoning.

---

<sup>47</sup> It is important to note that these games are still to be considered ‘one shot’ games – since no results of the interactions in the pre-test are given until after any post-test is concluded.

The second and third potential problems once again go back to the likelihood of a ceiling effect within the experiment. Except, rather than the framing effect as the product of the games (and thus to a certain extent minimisable by the researcher), they concern the types of individuals participating in the study, as well as the location in which the investigation is conducted. Deliberative mechanisms such as citizens' juries usually select their participants according to two principles: representation and randomisation. The first, by using some form of stratified sampling, is necessary to ensure that the panel represents a good cross section of the population affected by the issue under consideration. The second, randomising the selection of individuals within these groups by using something like electoral roll data, further guarantees the credibility of any policy recommendations that the jury may make. However, irrespective of the fact that many of these forums often offer a small honorarium for participation, there is still a choice to give up leisure (or professional) time in order to attend them. Demographically, the forums might indeed be diverse; but the self-selection bias towards those with a naturally more pro-social perspective, willing (even eager) to participate in a collaborative decision making process, poses certain possible problems. There is, for instance, a possibility that all the individuals involved might already employ team reasoning during the pre-test phase. If this is true, it becomes impossible to investigate whether deliberation may trigger the specific revision from an 'I' to a 'we' perspective. Moreover, since (for logistical reasons) both the pre and post-test stages are to be conducted at the site of the forum, it might also be the case that the physical surroundings – actually being at the site of the deliberation – might also yield the same problematic result.

To a certain extent, these problems are rather inherent to the project, and would require significant financial resources to entirely negate them: large payments would have to be offered to attend the forums, as well as the partaking individuals being asked to complete the pre-test before attending. Without such resources, the most obvious way to circumvent these ceiling effects (other than in the game design itself) is to therefore construct the experimental administration in a manner so it primes individual, self-interested behaviour to counter the more pro-social influence that person-type and situation-type bias may involve. As long as the same context is provided in both the pre-test and post-test phases, then it is still the effect of the deliberative mechanism that is being measured – it remains a *ceteris paribus* analysis.

There is also a potential difficulty with the monetary incentives used as payoffs in the game interactions. For starters, the debate introduces the question of whether actual payoffs are needed, or if hypothetical values are sufficient for the task (such as those used in some of the early work in experimental economics, see Thurstone 1931). The single most persuasive reason for adopting the latter, then, is undeniably a matter of ease of application. After examining a series of identical experiments conducted with both real and hypothetical payoffs, for instance, Thaler (1987) found that there was in fact not a significant difference between the two sets of results for particular types of experiments. Yet, in some reported cases (as far back as Smith 1962), there *has* been divergence between results based on the two different approaches. Wallace and Friedman (1942; pp. 179-180), in particular, provided the seminal economist rooted attack on the hypothetical approach, arguing:

For a satisfactory experiment, it is essential that the subject give actual reasons to actual stimuli [...] Questionnaires or other devices based on conjectural responses to hypothetical stimuli do not satisfy this requirement. The responses are valueless because the subject cannot know how he would react.

(Wallace and Friedman 1942; p. 180)

The central point of conflict between hypothetical and real payoffs, then, does not rest on a logical assertion that the former is in any way better than the latter. Instead, it is an issue of sufficiency: do individuals treat hypothetical outcomes the same as real ones when considering a decision they face? In experiments that use a large number of participants, and where individuals are required to make repeated choices during lots of different games, it is easy to see why using imagined payoffs might be preferable. But in the context of this research project, where only a small number of individuals are required, there appears little reason as to why some monetary rewards cannot be offered.

Finally then, a particularly strong argument levelled against much experimental work in economics is the claim that certain games are simply too complex for individuals not trained in a social science discipline to be able to fully comprehend. This is especially true for games with a number of different alternative strategies, where confusion regarding the impact that specific choices might have can cloud a rational calculation

made by the players. Two-person games with three choices are, to some degree, susceptible to this problem (or at the least more susceptible than games with less players and choices). If individuals are unable to make meaningful choices based on rational calculation, then the link between preferences/modes of reasoning and human action is broken. If this link is broken, then the experimental approach becomes unsuitable. To minimise the potential for this to happen, one of two courses of action are usually taken: either individuals are allowed to participate in a number of practice games before the real ones are attempted, or that more simple games are used initially (with meaningful data still generated) in order for players to ‘learn’ how to play them. In relation to this project, a mixture of both techniques to minimise the potential for this problem can be applied. There is no need to go beyond games with two strategy choices per individual, and practice encounters can be offered for players to become familiar with the set up of the experiments.

To summarise very briefly, up to this point in the chapter I have discussed a number of issues regarding some of the general methodological considerations that must be taken into account in this thesis. I have demonstrated that an experimental economic approach to the question of whether deliberation revises agency within individuals, inducing them to team reason, is both an entirely appropriate and effective methodology to investigate such a hypothesis.

#### **4.4 CONSTRUCTING GAME THEORETIC ENCOUNTERS**

---

To restate some points made earlier in this chapter, each interaction must have the following properties if it is to satisfy the internal validity requirement that certain strategies are indicative of certain preferences and, most importantly, certain modes of reasoning:

- i. There must be a single Nash equilibrium on the basis of individual, self-interested preferences.
- ii. The strategy choice that yields (i) must also be a dominant strategy for both players.
- iii. A collectively rational outcome that is the product of another set of strategies, that is not a Nash equilibrium in its own right.
- iv. The interaction must be non-decomposable.



With these four distinct requirements in place, five different games were constructed. Each was then also subjected to basic preliminary testing (asking ten random University of York students to play them) in order to identify whether any were susceptible to the ceiling effect discussed above. The five games that were chosen, and the associated preliminary test results, are now detailed.

#### 4.4.1 FIVE DIFFERENT GAMES

		Individual $j$	
		A	B
Individual $i$	A	5 , 5	10 , 4
	B	3 , 10	7 , 7

Figure 4.6: Game Theoretic Encounter One

In this game, there is a Nash equilibrium at [A, A]: if individual  $i$  plays strategy A, then individual  $j$  should follow the same course of action, since:

$$u(A) = u(5)$$

$$u(B) = u(4)$$

It is also the case that A is a dominant strategy for player  $i$ :

If player  $j$  plays A, player  $i$  should play A since 5 is greater than 3

If player  $j$  plays B, player  $i$  should play A since 10 is greater than 7

There is also a collective rational outcome at [B , B]. However, it cannot be sustained as a Nash equilibrium with other-regarding preferences because the game cannot be decomposed into independent decision vectors: the differences in payoffs between strategies A and B for player  $i$  are  $\{-2\}$  when  $j$  plays A, and  $\{-3\}$  when  $j$  plays B.

A preliminary test showed 30% of individuals played the team reason strategy automatically.

		Individual $j$	
		A	B
Individual $i$	A	7 , 7	2 , 11
	B	10 , 2	4 , 4

Figure 4.7: Game Theoretic Encounter Two

In this game, there is a Nash equilibrium at [B, B]: if individual  $i$  plays strategy B, then individual  $j$  should follow the same course of action, since:

$$u(A) = u(2)$$

$$u(B) = u(4)$$

It is also the case that B is a dominant strategy for player  $i$ :

If player  $j$  plays A, player  $i$  should play B since 10 is greater than 7

If player  $j$  plays B, player  $i$  should play B since 4 is greater than 2

There is also a collective rational outcome at [A , A]. However, it cannot be sustained as a Nash equilibrium with other-regarding preferences because the game cannot be decomposed into independent decision vectors: the differences in payoffs between strategies A and B for player  $i$  are  $\{+3\}$  when  $j$  plays A, and  $\{+2\}$  when  $j$  plays B.

A preliminary test showed 30% of individuals played the team reason strategy automatically.

		Individual $j$	
		A	B
Individual $i$	A	10 , 10	20 , 2
	B	2 , 16	15 , 15

Figure 4.8: Game Theoretic Encounter Three

In this game, there is a Nash equilibrium at [A, A]: if individual  $i$  plays strategy A, then individual  $j$  should follow the same course of action, since:

$$u(A) = u(10)$$

$$u(B) = u(2)$$

It is also the case that A is a dominant strategy for player *i*:

If player *j* plays A, player *i* should play A since 10 is greater than 2

If player *j* plays B, player *i* should play A since 20 is greater than 15

There is also a collective rational outcome at [B , B]. However, it cannot be sustained as a Nash equilibrium with other-regarding preferences because the game cannot be decomposed into independent decision vectors: the differences in payoffs between strategies A and B for player *i* are  $\{-8\}$  when *j* plays A, and  $\{-5\}$  when *j* plays B.

A preliminary test showed 20% of individuals played the team reason strategy automatically.

		Individual <i>j</i>	
		A	B
Individual <i>i</i>	A	11 , 11	2 , 16
	B	17 , 3	4 , 4

Figure 4.9: Game Theoretic Encounter Four

In this game, there is a Nash equilibrium at [B, B]: if individual *i* plays strategy B, then individual *j* should follow the same course of action, since:

$$u(A) = u(3)$$

$$u(B) = u(4)$$

It is also the case that B is a dominant strategy for player *i*:

If player *j* plays A, player *i* should play B since 17 is greater than 11

If player *j* plays B, player *i* should play B since 4 is greater than 2

There is also a collective rational outcome at [A , A]. However, it cannot be sustained as a Nash equilibrium with other-regarding preferences because the game cannot be

decomposed into independent decision vectors: the differences in payoffs between strategies A and B for player  $i$  are  $\{+6\}$  when  $j$  plays A, and  $\{+2\}$  when  $j$  plays B.

A preliminary test showed 30% of individuals played the team reason strategy automatically.

		Individual $j$	
		A	B
Individual $i$	A	4 , 4	20 , 2
	B	3 , 20	13 , 13

Figure 4.10: Game Theoretic Encounter Five

In this game, there is a Nash equilibrium at  $[A, A]$ : if individual  $i$  plays strategy A, then individual  $j$  should follow the same course of action, since:

$$u(A) = u(4)$$

$$u(B) = u(2)$$

It is also the case that A is a dominant strategy for player  $i$ :

If player  $j$  plays A, player  $i$  should play A since 4 is greater than 3

If player  $j$  plays B, player  $i$  should play A since 20 is greater than 13

There is also a collective rational outcome at  $[B, B]$ . However, it cannot be sustained as a Nash equilibrium with other-regarding preferences because the game cannot be decomposed into independent decision vectors: the differences in payoffs between strategies A and B for player  $i$  are  $\{-1\}$  when  $j$  plays A, and  $\{-7\}$  when  $j$  plays B.

A preliminary test showed 20% of individuals played the team reason strategy automatically.

#### 4.4.2 ALLOCATING CO-PLAYERS: DEFINING THE ‘WE’

Before I discuss the exact process of how the tests were administered, I need to discuss an important point regarding the allocation of co-players. To make sure the games represented one-shot encounters, and therefore strategy choices could not be affected by

repetition, allocation of co-players was anonymous and randomised. However, when studying a phenomenon that is directly related to social identity, the *group* from which each co-player  $j$  is chosen will clearly have a significant impact on the strategy choices of the individual  $i$ . In this case, the most obvious level at which a social identity can exist is within the deliberating group itself. On this basis, an individual who team reasons when participating in such an encounter has done so because that particular social identity has been made most salient. This is the hypothesis that was developed in chapter two, and expanded upon in chapter three.

However as I also made clear, there are alternative groups that social identity might be formed on the basis of. Deliberation might indeed trigger team reasoning in relation to members of the deliberating group, but it also might have an impact beyond this. It is possible that by priming one specific social identity, others may also become more salient for a given individual. For example, one possibility is to broaden what counts as the ‘we’ to include members of *any* deliberating group. On this account, social identity therefore refers to the group of individuals involved in a deliberation, but not necessarily with each other. If individual  $i$  chooses the team reason strategy when playing a game with an anonymous individual  $j$  from an alternative deliberation, then a social identity amongst ‘deliberators in general’ has clearly emerged. Similarly, it is also possible to broaden the definition of the ‘we’ to the most universal level: the general public. In this case, the experiments will show the extent to which an identity (arguably) similar to that of a citizen has been created. A shift from individual to team reasoning would be indicative of a process that has primed a sense of common feeling amongst all individuals irrespective of their participation in the deliberation.

There are then, numerous other prospective definitions of a ‘we’ that might be investigated in the process. Yet the presence of team reasoning within a deliberating group remains the fundamental proposition to examine, as it represents the simplest level at which deliberation can impact upon social identity. Given this result, the first three games are devoted to testing the presence of team reasoning within the deliberating group, the fourth in relation to a member of the general public, and the fifth in relation to a member of *any* deliberating group. The choice of these other two forms of ‘we’ will be more comprehensively justified and discussed in chapters six and seven of the thesis.

#### 4.4.3 INSTRUCTIONS AND EXPERIMENTAL ADMINISTRATION

The last point I want discuss in this section concerns the administration of the experiment, and how the games were both presented to, and played by, the individuals involved. Following convention in experimental economics, and to ensure both standardisation across different participants as well as different stages, identical and clear instructions were given at each point in the process. Immediately before any individuals were presented with the games, the following verbal information was given:

You are going to make several decisions, from which you can earn points that will be converted into lottery tickets to win two prizes of £100.

You are will be paired up with a second, anonymous individual, and will be presented with the identical problem. You will then be asked to choose one of two courses of action, either A or B. To work out the likely consequences of any decision, you will have to take into account what the other person is likely to choose.

Please be aware that in games one, two and three, you will be partnered with an anonymous member of this deliberating group. In game four you will be partnered with an anonymous member of the general public. In game five you will be partnered with an anonymous member of a second deliberating group.

Your decisions are completely anonymous, to both your co-player and the researcher. Please also be aware that your choices in one game have no impact on the choices you make in another. Each game is entirely independent.

Finally, participants were then given the five games to play at various points (which I will outline in more depth during the next chapter). A copy of the experiment layout, and instruction sheet, is given in appendix one.

#### 4.5 CONCLUDING REMARKS

---

In this fourth chapter, I have outlined both *why* experimental economics provides an effective technique to investigate the presence of team reasoning, and moreover, why it is appropriate in combination with studying political deliberation. Having set out the case for the experimental side of the methodology, I now progress to how it is applied.

Or in other words, in the next chapter I outline the structure of the deliberation, in this case a mini public, which was used to empirically test the hypothesis regarding agency revision and team reasoning.

## — CHAPTER 5 —

# THE CASE STUDY OF DELIBERATION: A PEOPLE'S INQUIRY INTO PERSONAL INFORMATION

### 5.1 INTRODUCTION

---

During the first chapter of this thesis, I discussed the concept of deliberation, and its relationship with democratic theory. I then used the rational choice approach to distinguish between the three layers at which deliberative revision was possible, offering an account of some of the empirical work currently being undertaken in this field. I then focussed on the third layer, namely agency, and argued that deliberation has the potential to induce individuals to reason on behalf of a team – drawing largely on a conceptual framework developed in social psychology. In chapter four, I then outlined a specific methodology that can be used to measure such a shift, before detailing the setup of the experimental side of the investigation.

To complete this account of the methodology, then, the next step is to provide the details of the specific deliberation that was used as a case study for the investigation.

As such, my three concerns in this chapter are as follows:

- i. To provide a brief outline of the various attempts in deliberative theory to create different forums used to approximate the ideal conditions of deliberation.
- ii. To identify the key criteria that help ensure high quality deliberation in practice.
- iii. To provide an outline of the actual deliberation used to investigate the presence of team reasoning amongst individuals, and analyse it with reference to the key criteria identified above.

Once both these objectives have been achieved, I can then progress to the final sections of the thesis: an analysis, and then application of the results back to deliberative democratic theory.



## 5.2 DELIBERATION IN ACTION

---

In an attempt to institutionalise political deliberation, and deliver the benefits that normative theory so suggests, a number of different types of deliberative forums have been developed. Commonly referred to under the general heading of ‘mini publics’ (Goodin and Dryzek 2006), these different initiatives have included citizens’ juries, assemblies, consensus conferences and even Deliberative Polls. On one side, they have often been used in order to inject increased levels of legitimacy and democracy into decision-making. On another, though, empirical deliberative theorists have also used them to study the effects of deliberation on the individuals involved. It is this latter endeavour that proves relevant here.

In this chapter I want to start by outlining some of the more common ways in which deliberation has been institutionalised in these real life settings. In doing so, my objective is to provide the context for the identification of a number of *empirical criteria* (see figure 1.3 in chapter one) relevant for securing high quality deliberation in practice. But before I do this, and following in the footsteps of Goodin and Dryzek (2006), one of the most commonly discussed examples of a deliberative process needs to be considered and then dismissed. Crucially, in doing so, it generates the first empirical criterion relevant for the construction of a successful deliberative forum.

Participatory budgeting in Porte Alegre (Brazil) undoubtedly represents one of the first incursions of a formal deliberative process into public policy, where from 1989, massive inequality motivated the city to take drastic action on the manner in which public funding was allocated. Split into three different levels, participatory budgeting involves Regional Assemblies (open to the general public), who then elect members to serve on the Regional Budget Forums, who then in turn elect members onto the decision-making Municipal Budget Councils. Clearly, this process has many features associated with deliberative models of decision-making – most obviously the involvement of ordinary citizens who are asked to debate and discuss policy decisions. But whilst it may fulfil many of the obligations to claim deliberative status, it also fails in a bid for the label of being a genuine mini public for one distinct reason. The participants involved in the process are either self-selected in as much as *anyone* can attend a Regional Assembly, or in the case of the latter two levels, only if they are

elected. What makes the mini public model so different, as is claimed in both Smith (2009) and Goodin and Dryzek (2006), is the selection process for participants employed by those running the forum. All mini publics rely on some form of random selection<sup>48</sup> to obtain their participants – although the exact manner in which this takes place can differ from one forum to another. Whilst I will discuss the technicalities of the mechanism of randomised choice a little later, with respect to the concept of representativeness, for now, though, one empirical criterion can be established: the random selection of participants.

*Citizens' Assembly: British Columbia, Canada (2004)*

Held on the topic of electoral reform, and specifically the proposal for an alternative voting method, the British Columbia case study provides an almost paradigmatic example of how an effective and well-conceived deliberative process can be run (Lang 2007, Warren and Pearse 2008). It involved 160 randomly recruited citizens (one male, one female from each province), who were brought together on various weekends over the course of a year. The first four months in Vancouver were spent largely on educative matters, as members learned the various intricacies and features of different electoral systems, the electoral history of the province, and general demographic information about the changing population.

Once this phase had been completed, a preliminary statement was issued on the subject, and the citizens then spent around two months travelling around the province, taking evidence from a wide variety of stakeholders including (conflicting) interest groups, policy makers and the general public. Further to this, over 1600 written submissions were made to the assembly, with members taking time over the summer period to read and reflect upon them. Finally, 10 months after the process was first set in motion; the assembly spent six weekends back in Vancouver undertaking the final stages of the official deliberation. Debating the advantages of various electoral systems with respect to a set of criteria decided upon by the assembly itself (fairness/proportionality, voter choice and effective local representation), the final outcome of the process included a comprehensive report with associated recommendations, and a binding referendum held on the following question:

---

<sup>48</sup> The idea of random selection forming a crucial part of a democratic process, of course, is nothing new. As Held (2004) points out, even ancient Athens had particular offices of state that were filled on the basis of probability and chance, rather than democracy.

“Should British Columbia change to the BC-STV electoral system as recommended by the Citizens’ Assembly on Electoral reform?”

With significant media coverage surrounding the origin of the referendum the motion passed with 58% of the vote (winning in 77 of 79 constituencies), though unfortunately for most involved in the assembly, was just shy of the 60% super majority required.

*Citizens’ Juries: Health Policy, UK (1996)*

Citizens’ Juries, on the other hand, are much smaller in scale than the Assembly example of British Columbia. Originally conceived and then popularised in the 1980s by the independent American based Jefferson Centre, it wasn’t until the late 1990s, particularly driven by the impending general election and subsequent victory of Tony Blair’s Labour Party, that this model found its way across the Atlantic to the UK (Parkinson 2006). Usually requiring 12-24 randomly selected citizens, the jury format involves lots of small-scale facilitated deliberation; evidence from expert witnesses with the power to question/recall them; and the publication of a report detailing the findings and decision of the group.

The example discussed in Coote and Lenaghan (1997) then, describes five separate juries on the general topic of health care. Each jury was given their own specific sub question to consider, ranging from how the NHS should be funded, to the provision of mental health services in a local area. Individuals were recruited by an external commercial organisation using a stratified methodology, based on a ‘profile of the local population derived from census and other data’ (1997; p.9), and paid £200 each. The five juries were then brought together for four days, with the IPPR compiling the agenda, as well as ultimately in charge of writing the final report. Finally, two trained facilitators were employed throughout the four days.

*Consensus Conference: Telecommunications Industry, USA (1999)*

The third common form of mini public, the consensus conference, originates from the Danish Board of Technology in the late 1980s, and to date over 60 have been run across the globe (Hendriks 2005). Similar in many respects to a Citizens’ Jury, the original incarnation utilised 15 individuals (split into three groups of five), who initially met for two weekends to plan the four-day deliberative forum.

For example, in 1999 a number of US partner institutions including the Education for Public Inquiry and International Citizenship (EPIIC) programme at Tufts University, *Technology Review* Magazine at MIT, the College of Social and Behavioural Science at UMASS, and the National Science Foundation, came together to organise the first US incarnation of this form of mini public. Held on the topic of telecommunications, and more specifically the question of universal provision of Internet access, the organisers recruited seven men and seven women respectively, who after the familiar weekends spent preparing for the event, then met for three days worth of discussion and deliberation. On the final morning, a provisional four page ‘consensus statement’ was issued at a press conference.

*Deliberative Poll: Power 2010, UK (2010)*

Originally and most often associated with James Fishkin and the Stanford based *Centre for Deliberative Polling*, Deliberative Polls ® have become some of the most high profile examples of mini publics in action. The fundamental design includes the collection of up to 500 random citizens, who are given a preliminary survey to complete on the topic of the forum. Held over one to two days, the members are further split into smaller groups of around 15-20 individuals, each led by a trained facilitator, and are asked to discuss/debate with each other, as well as come up with questions for various expert witnesses made available to them. The event culminates in a second completion of the original survey, to test any change in attitudes that have occurred over the course of the deliberation. Crucially, no binding decision (in any format) is required in the process, other than the individual questionnaire submissions.

The Power 2010 poll then, as an example, collected a sample of 130 citizens from around the UK, on a demographic representative basis according to gender, age, marital status, party affiliation, income, ethnicity, employment and region inhabited; as well as a number of other more nuanced factors such as interest in current affairs (very narrowly defined by purchase of daily newspaper, consumption of news media etc). A significant amount of balanced pre release material was then sent out to each participant on the topic of political reform, including information on the current state of UK democracy and the various proposals being discussed in the public sphere. Finally, participants were then brought together at a large venue in London, and spent two days split into smaller groups of 10-12 individuals, discussing the various proposals before

‘distilling the many ideas’ they received into a manageable shortlist. This shortlist has now been published, and forms the core campaigning principles of the movement.

Let me summarise this section very briefly then. By outlining some of the more common ways in which the theory of deliberation has been put into practice, I have touched upon some key operational features common across the different types of forum. Relating this to the research question of this project, the relevant issue to consider is the impact these features have on determining the quality of deliberation, and therefore, the extent to which any forum is truly deliberative in the normative political sense discussed in chapter one. It is this point that I now take up, firstly identifying and secondly discussing the key empirical criteria emanating from these examples. Once I have done this, I will then be able to normatively ‘judge’ the case study used for the agency revision investigation.

### **5.3 EMPIRICAL CRITERIA FOR DELIBERATION**

---

From examining these four different types of mini publics, a number of empirical criteria that directly affect the quality of deliberation can be identified, which I now discuss in turn. These will then be used as a framework to evaluate the particular forum chosen for the study.

#### **5.3.1 RANDOM AND REPRESENTATIVE SELECTION OF PARTICIPANTS**

Already established as the first criterion, the notion of random selection is a stated requirement for all of the different mini publics. Where individuals are required to make decisions, on any topic, the makeup of the group will clearly be pivotal to the final result. Deliberative projects, at their very epicentre, are premised on the prioritisation of citizenship and the involvement of the public in democratic decision-making. Due, in large part, to the nature of how mini publics have been used as a proxy for this, it is clear that without some form of selection criteria for the participants, it is both possible and even probable, that the makeup of the assembly, jury, conference or poll will form victim to some kind of bias. In particular, where the topic under consideration is controversial, and participation in the deliberation open to the general public (as in the case of Porto Allegre), then we are likely to see those with a raised level of interest in the outcome, those who simply love participating, or indeed those with a particularly controversial perspective on the issue getting involved in

disproportionate numbers. Conversely, we are also unlikely to see those individuals who might declare themselves as uninterested or unfamiliar with the topic; or even those who might consider such forums as intimidating.

The element of this criterion, representative selection, is linked to the first in that it is also concerned with the make up of the deliberative forum. For any mini public to claim legitimacy – either in the case of directly affecting public policy, or by making a recommendation to the general public – then the demos must respect the outcome of the process as being valid. Clearly, this is directly related to the individuals involved in the process. If it does not reflect the population affected by the decision, then its legitimacy can be called into question. The question, of course, becomes in what sense, or on what level should this comparison be made? In aggregative accounts of decision-making, it is clear that the most important dimension is via preferences or values. If all that counts is the representative's vote, then it is obvious that individuals will feel most legitimately represented by a forum that includes participants who cast their ballot in the same manner as they would themselves. Extending this to mini publics, the logical deduction would therefore be to ensure the makeup of raw preferences of the participants matches the makeup of raw preferences of the population as a whole. But this is exactly where the difference with deliberative models of decision-making lies. Mini publics are dynamic, where a fundamental assumption is made that preferences post deliberation are more important, or carry more weight, than those that are raw. Because of this, the way in which an individual both acts and changes over the course of the deliberation is of equal importance as the preferences they initially hold. In other words, deliberative representation becomes a much more complex notion, one which needs to consider a host of other descriptive features of the participants.

Both these issues are particularly relevant when considering how the makeup of the deliberative forum impacts upon the quality of the deliberation found within it. In circumstances where it is disproportionately dominated by individuals already involved in an issue, or indeed by those who might attend to pursue a particular agenda, then a number of the behavioural criteria established in chapter one are much more difficult to obtain. For example, if the forum makeup consists wholly of people already committed to a specific policy or perspective, then the maxim of mutual respect and reciprocity is more likely to be relaxed. Individuals are more likely to listen to, but then ignore, other people's points in a discussion. Additionally, the public principle is more likely to find

itself relaxed, as individuals already committed to specific positions find it more difficult to take a new, public perspective. By randomising the representative selection of the participants, these potential problems are guarded against, and thus higher quality deliberation in the forum is promoted.

### 5.3.2 SMALL GROUP DISCUSSION

Fundamental to all of the mini publics discussed above is the focus put on small-scale deliberation. Even in the case of the two larger forums, Deliberative Polls and Citizens' Assemblies respectively, smaller subgroups are consistently utilised for discussion during the course of the overall process. There are a number of key reasons for this, all of which relate directly back to nature and quality of the discursive process. Firstly, as I have already discussed in chapter one, deliberation is an extremely time consuming endeavour. How effective a discussion on a complex topic is, which includes comprehension and consideration of all points of view, challenging evidence and assumptions, and then culminating in a decision-making phase, is linked to the number of individuals involved in the process. By this, I mean there is a strong positive relationship between the time that is required for such a process to take place, and the amount of people participating in the actual discussion. The greater the number of perspectives that is included in the process, then the longer, by definition, it will take.

Moreover deliberation, as articulated in chapter one, is a very particular form of speech. It is not simply 'talk in action' (Heritage and Clayman 2010), and is characterised by the various behavioural criteria established earlier. One of which is the maxim of equality. According to this criterion, deliberation should include ample opportunity for all members to contribute equally to the discussion. Since most, if not all, mini publics have defined timescales that are often dictated by factors such as cost, it is imperative to ensure that the quality of deliberation is not compromised. Breaking larger groups down into smaller subgroups throughout the course of the process is essential if the maxim of representative selection, is to be satisfied at the same time as guaranteeing that deliberation, in this technical sense, can take place. In short, reducing the size of the deliberating group at particular points in the wider process allows higher quality deliberation to take place, without wholly sacrificing the legitimacy of the overall forum by restricting the number of different inputs into it.<sup>49</sup>

---

<sup>49</sup> This, of course, raises legitimacy questions, as the restriction of inputs can have a significant effect on the outcome of the deliberation. I will consider this particular point in chapter seven, when I discuss the

Finally, it is also worth pointing out that deliberating in small groups is much less intimidating for individuals to take part in, where fears over public speaking might prevent some members from offering opinions and challenging points of view. This is particularly relevant for those who hold minority perspectives. Deliberation should value ideas and opinions on the basis of their content (the public principle), rather than the number of participants who come into the process holding them. For example, in cases where a particular perspective performs well against deliberative criteria but is held only by a single individual, then it is plausible to imagine a situation where it might not be put forward for fear of contradicting a large majority. By reducing the size of the discussion group, though, more equality in participation should be secured as individuals face much less pressure when contributing in the deliberation. In short, small groups help to ensure higher quality deliberation on this dimension.

### 5.3.3 FACILITATED DISCUSSION

Although not directly present in the normative statements on deliberation, the use of a facilitator during mini publics is almost universally accepted as a requirement for an effective deliberation to take place. There are a number of crucial reasons for this, but all rely on the proposition that left to their own devices, a group of individuals cannot be guaranteed to both self manage, and participate in, a deliberative process that would compare favourably to the normative ideals already discussed. Firstly, there are clear organisational/administrative benefits from having an individual involved in the group, but playing no ‘substantive’ role in the deliberation. Having time to focus on making sure deadlines for decisions are met (acting as time keeper), that all topics are considered, and that evidence/information can be requested throughout the process, are all essential for the effective functioning of any mini public. The facilitator in each group, more often than not, fulfils these operational roles.

Secondly, whilst utilising small group deliberation allows for the capacity of equality to be realised – there is enough time for all individuals to participate in the discussion – it does not, however, ensure this maxim is satisfied. Capacity and actuality are two very different notions, and although there might be enough time for every individual to contribute, it does not necessarily follow that this will happen. Deliberation on a

---

shift from second to third generation models of deliberative democracy in relation to the experimental results.



controversial topic, with participants who are directly affected by the outcome and therefore have a personal stake in the group's decision, can be easily hijacked or manipulated by single participants. Shaping the discussion by talking more often or even louder than other individuals, or forming strategic alliances with other members to dictate the trajectory of a discussion, are both possibilities when there is no independent third party to moderate or safeguard the process.<sup>50</sup>

Extending beyond this, because mini publics should be made up of individuals from a wide variety of different social circumstances and with a varying degree of discursive capacity (reflecting the demos), the role of the facilitator is also just that – to facilitate the discussion. Prompting individuals to listen actively, helping them to uncover shared vocabularies (Miller and Rose 2008) so they understand each other's points of view more fully, and ultimately, setting the tone of the deliberation as one of mutual cooperation, are just three of the possible tasks that good facilitation demands. In short, whilst the concern of those partaking in the deliberation is the topic being considered, the facilitator focuses only on the nature of the discussion. In this sense, their role is to ensure both the behavioural and structural criteria of deliberation (identified in chapter one) are fulfilled.

#### 5.3.4 PRE-RELEASED MATERIAL

One of the fundamental normative principles that underpin deliberation, that it helps individuals or groups to make 'better' decisions, is premised on the idea of education (Arendt 1970). In fact, the argument that debate and discussion is required for humans to make rational choices goes as far back as Mill, who famously defended the right to free speech on the grounds of fallibility. It should be no surprise, then, that the most common topics chosen for mini publics are those that on first glance, appear quite complex and require a certain degree of specialist knowledge to comprehend. Subjects such as electoral reform, which demands at least a basic understanding of how various mathematical social choice mechanisms work, or health care reform, which demands an understanding of both medical and governmental practices and institutions, are two good examples. Deliberative preferences are assumed to be not only 'better' according to the public principle, but also on educational grounds.

---

<sup>50</sup> Sunstein's (2002) 'Law of Group Polarization' is a particularly nice possibility of what can happen when deliberation goes wrong, and is hijacked by individuals from the extreme ends of the policy spectrum.

Clearly, within the deliberative processes of evidence submission, discussion and debate, there is a large scope for education. Not simply learning what others think about an issue, or how a particular policy might affect a particular group of individuals, but rather learning in the purer sense of the term. In the case of British Columbia for example, before they could take part in any debate on the merits of a particular voting system, the participants first had to understand exactly how they all functioned. The differences between systems such as alternative vote, alternative vote plus or single transferrable vote are extremely subtle, and require an understanding of the rather formal distinction between preferential and proportional outcomes. Whilst presentations from defined 'experts' on topics like this are used to inform/educate the members during the mini public itself, it is done so, equally as often, in conjunction with pre-released material.

Sending out detailed information, whether in written, audio or multimedia format, before individuals arrive at the actual mini public has a number of positive externalities. Firstly, where there is a requirement to convey simple, more factual data (such as figures of electoral turnout or break down of voting patterns etc), then it is obvious that time is better spent understanding and digesting this before individuals arrive at the forum. Time spent sitting and reading, is time not spent questioning, discussing and debating. Secondly, pre released material helps to minimise the 'expert effect' within the small group discussion. By giving all members some basic knowledge of the topic before they arrive, the objective is to remove the distortive impact that an individual with a little familiarity with the topic might have, preventing them from dominating the discussion under the auspices of being a self-nominated expert. And thirdly, it helps to prepare the individuals for the ethos and atmosphere of the mini public they are attending. Participants are much more likely to arrive with the mindset of an individual taking part as a responsible citizen, rather than a consumer attending a focus group. On the basis of these points, the use of pre-released material therefore has a distinct positive impact upon the quality of the deliberation found within a forum. On one level by priming individuals before they attend a forum, it helps to ensure that the various behavioural criteria of good deliberation are satisfied. And on a second level, it also plays a significant role in the first phase of the structural account of deliberation, educating the individuals involved in the process on the topic being discussed.

### 5.3.5 OUTCOME INFLUENCE

The last empirical feature that I want to discuss also relates to a particular phase of the structural account of deliberation. This time, though, it concerns the final point of the process, where participants are required to come to a collective decision. As I have already discussed in chapters one and three, the presence of this phase has a distinct impact on the nature of deliberation. But there is also a second aspect to this feature that has not been considered, one that concerns the *way* in which the decision reached is then used. To put it most simply, forums that are convened on the basis that the decision will be put into practice are likely to operate differently from those where no consequence is attached to the collective decision. Indeed, it is quite true of everyday life that individuals moderate and alter their behaviour when interacting in particularly important circumstances. Goodin (2008; pp. 19-36), discusses a number of different ways in which mini publics have traditionally impacted the policy process:

- i. Actually making policy
- ii. Being taken up in the policy process
- iii. Informing public debates
- iv. Shaping policy by market testing
- v. Legitimizing policy
- vi. Confidence building
- vii. Popular oversight
- viii. Resisting Co-option

Normatively speaking, there are clearly many reasons *why* it is preferable to involve more citizens directly in policy-making processes. But that still does not really give a sense of a) how this might affect the nature of the deliberation, or b) how this might then be categorised according to some kind of scale. I suggest all these different examples of impact can be analysed according to how they answer two different questions. Firstly, is the decision of the mini public binding? Clearly, in cases where the outcome of the deliberation must be accepted by the commissioning body, there is a heightened sense of importance aligned with the decision-making phase. If participants are aware that the actual policy or proposal they recommend is the one the organisation must take on, then a likely consequence is for them to take the process much more seriously, and ultimately, to partake in the deliberation more fully. This, in turn, has a direct impact on the nature and quality of the deliberation. If individuals take the

process more seriously, and see their role as central to the process, then it is plausible to suggest they become more likely to act in accordance with the behavioural criteria established earlier. This is particularly true if the ‘rules’ (or norms) are made explicit early on in the process. Secondly, the other question that must be answered concerns the power of the commissioning body to then act upon the decision that has been made by the forum. In cases where the body has the ability to directly translate decision into policy, then the outcome of the deliberation is made even more significant. In this example, individuals partaking in the forum will be much more aware that their contributions will have a direct effect on the lives of individuals linked to the issue being considered. Again, this will re enforce the seriousness of the forum, and ultimately help to ensure higher quality deliberation.

## **5.4 THE CASE STUDY: PERSONAL AND PRIVATE INFORMATION**

---

With these empirical criteria established, and more importantly, the manner in which they might affect the quality or nature of deliberation discussed, the next step is to offer an outline of the forum used for the experimental investigation. Once I have done this, I will then apply the former to the latter, to demonstrate the appropriateness of this particular mini public for investigating political deliberation in practice.

### **5.4.1 THE COMMISSIONING BODY AND TOPIC**

The mini public chosen for the case study was a process commissioned by the London-based think tank Demos, in collaboration with both the Information Commissioner’s Office (ICO), and the organisation Consumer Focus. It was held on ‘Private and Personal Information’, a topic which had generated significant amount of media attention since an incident in 2007, where the National Audit Office and HMRC lost two discs containing the personal information (including NI numbers, bank account details and addresses) of 25 million individuals and 7.25 million families. The stated objectives of the project were:

- i. To obtain a considered insight into what people really think about how personal information is (mis)used, and;
- ii. To formulate some appropriate controls and governance structures, that people think should be applied to this industry.

#### 5.4.2 RECRUITMENT OF PARTICIPANTS

Due to the nature of the topic, familiarity with technology and in particular the amount of interaction with online services was recognised very early as a likely strong influencing factor on the debate. Because of this, the decision was made to run two forums concurrently. One based in London, with residents from all over the capital, and the other based in Bradford. By convening forums in cities with the highest (London), and lowest (Bradford) home internet penetration rates in the UK, the idea was to additionally investigate the extent to which this variable in particular might affect the deliberation, and consequent outcome of the process. The participants for both forums, then, were recruited by Criteria, an external firm specialising in UK-wide social science fieldwork and market research.

#### 5.4.3 DEMOGRAPHIC OF PARTICIPANTS

The demographic break down of the participants in both forums was as follows.

By sex:

---

Table 5.1: Sex of Participants

---

Location	Male	Female
London	63.2%	36.8%
Bradford	61.1%	38.9%

---

By age:

---

Table 5.2: Age of Participants

---

Location	18 – 30 years	31 – 40 years	41 – 50 years
London	36.8%	36.8%	26.4%
Bradford	38.9%	33.3%	27.8%

---

By ethnicity:

---

Table 5.3: Ethnicity of Participants

---

Location	White British	British Asian	British Indian	British African	British Caribbean
----------	---------------	---------------	----------------	-----------------	-------------------

---

London	73.7%	5.3%	15.7%	5.3%	5.3%
Bradford	61.1%	38.9%	0%	0%	0%

By social class:

Table 5.4: NRS Social Grade of Participants<sup>51</sup>

Location	B	C1	C2	D
London	26.3%	36.8%	21.1%	15.8%
Bradford	16.7%	38.9%	11.1%	33.3%

#### 5.4.4 THE STRUCTURE OF THE DELIBERATION

Each forum, whether in London or Bradford, took place over the course of a month long period from October 21<sup>st</sup> and October 24<sup>th</sup> 2009 respectively. Participants attended four weekly sessions, lasting approximately four hours each, at the same venue on either Wednesday evenings or Saturday mornings. Both forums were randomly split into three discussion subgroups, comprising of six to seven individuals, each with a Demos facilitator. Each week, a modest amount of pre release material was given to all members of the group on the following week's subject, which included both factual and polemic pieces of text.

The general topic for each week was as follows, with the actual break down of the sessions available in Bradwell (2010):

Week 1: Privacy, consent and control

Week 2: Personal information online

Week 3: Personal health information

<sup>51</sup> A Little explanation of this variable is probably useful. The NRS grading system is a commonly used tool (in market research) to grade individuals according to a 'social scale'. The rankings represent the following descriptions, and are assigned via a combination of user response and recruiter evaluation:

Code	Social Class	Descriptor	% National Population (2008)
A	Upper Middle	Higher managerial or professional	4
B	Middle	Middle management or professional	23
C1	Lower Middle	Supervisory, clerical or administrative	29
C2	Skilled Working	Skilled manual work	21
D	Working	Semi-skilled and unskilled manual work	15
E	Lower	Unemployed, casual grade work	8

Source: NRS website (<http://www.nrs.co.uk/lifestyle.html>)

## Week 4: Regulation, safeguards and policy

There are a number of reasons as to why the overall topic was broken down into *these* four sessions. Most importantly, it represents a substantive escalation in terms of both complexity and specificity. Session one, for example, was designed to help the participants gain a general understanding of what constituted personal information, and moreover, to comprehend some of the legal definitions, possibilities and constraints upon it. The second session, however, then took this enhanced understanding, and got the participants to apply it to the most common source of controversy on the topic – online information. Week three took this a step further still, and focussed the group on a particular industry’s use of this data, that of the health sector. And finally, the fourth week looked at the future, and asked the participants to critique current policy before coming up with agreed proposals in light of this. The premise was for each session to act as part of the educative process for the next one, and ultimately, for the first three to provide the foundation for the final decision-making phase.

Indeed, this links quite nicely to perhaps the most distinctive departing feature of this mini public compared with those discussed above. Instead of holding the entire forum over a condensed time period of two to three days, the aim was to create a different type of deliberative event. One that took on board the common features discussed above, but also looked to innovate in order to address some of the potential issues that arise in the more common mini public models. In particular, the forum was designed to depart from the conventional mini public model on two dimensions, which I now briefly consider.

### *Separation of Deliberative Phases*

The most important element of difference, by far, was the desire to separate out the various stages within the deliberative process. Distinguishing between heavily educative, discursive and decision-making phases was considered, by the organisers, of paramount importance for the topic – in large part because of the complexity of both legislation and the current model of governance. This was particularly salient in sessions one and two, which had some significant technical concepts to understand and consider. The decision to focus more heavily on expert presentations and questioning during these first sessions was a direct attempt to nudge the participants away from discussion for persuasion, and onto discussion for education. It was, to put it quite

bluntly, an attempt to create a more independent and distinct learning phase of the deliberation.

After the initial weeks the focus of the sessions shifted to reflect the different stages in a deliberative process (reflecting the structural account of deliberation). Week three, in particular, asked the participants to debate the use of personal information in the provision of health care. Again, this topic was chosen because it both applied to all the participants, and secondly, because of the inherent personal nature of the issue. The objective was to promote as much discussion as possible, and moreover, for participants to become familiar, and even comfortable, with challenging each other's points of view. Again, to put it simply, it was an attempt to help participants learn how to deliberate, and to give them the confidence to debate ideas in their groups. And lastly, week four represented the climax of the event, where the topic of government regulation was introduced. The objective for this final week was for participants to come up with a number of agreed policy proposals that could be taken forward by Demos and the two other stakeholders. This session was designed to induce much more cooperative reasoning, and asked participants to draw on the previous three weeks of learning and debate, to decide as a group the courses of action they felt were appropriate.

To briefly conclude this point, by separating each section of the deliberative process out from each other, the aim was that individuals could (and would) focus more heavily on making sure they engaged fully in every session. For example, having an individual take part in the discursive phase without having taken part fully in the educative sessions is clearly not ideal. Or similarly, having the educative phase dominated by individuals jockeying for influence to exert during the decision-making stage also works to subvert the aim of that part of the process. Breaking the forum up into these defined subsections helped guard against this type of problem, and ultimately, was done to promote higher quality deliberation overall.

### *'Opening Up' Mini Publics*

As well as the substantive investigation into the subject matter, a secondary objective was to develop a methodology that allowed deliberative mini publics to be utilised more commonly by public policy making organisations. Deliberative forums that are organised to run over two days require a significant amount of time and resources to run. But they also require the participants to give up a significant amount of personal



time in order to attend them. Taking two days away from work, or even two days away from a family is simply not an option for many individuals, and because of this, their participation in deliberative forums becomes unlikely.

Opening up the process by breaking it down into these four distinct sessions (each lasting four hours long), the objective was to widen the scope for more individuals to attend. The drawback, clearly, was the potential for individuals to pick and choose the sessions they felt were more relevant or more interesting, or indeed it was possible for personal circumstances to prevent an individual from attending all four. But the use of a monetary reward for completing the process – including a bonus for attending all four sessions – worked almost perfectly to prevent this.<sup>52</sup> And indeed, what soon became apparent after the first week was the importance of camaraderie amongst the members, who after buying into the process very early on began to look forward to the following week's session.

## **5.5 THE CASE STUDY: HOW DOES IT COMPARE?**

---

Having established both a set of empirical criteria to analyse deliberative mini publics, and then given an overview of how my case study was constructed, the final step is to apply one to the other. How does the *People's Inquiry into Personal and Private Information* fare in terms of deliberative criteria specified in section 5.3. This final section of the chapter is crucial in determining whether the results of the experimental investigation can be seen as representative of high quality deliberation, and therefore, whether they can be used to draw any conclusions back to deliberative models of democracy.<sup>53</sup> Let me now discuss each criterion in relation to the case study.

### *Random and Representative Selection of Participants*

The selection of the participants was completed entirely at random, with each individual being paid £50 to attend the four sessions, and a £50 bonus for completion. None of the participants who attended had been involved in anything similar to this before, and indeed, none displayed a particularly strong opinion on the topic or methodology. The

---

<sup>52</sup> Indeed only one individual in the Bradford group, and none in the London group, failed to attend four sessions.

<sup>53</sup> This is, of course, only one way of 'measuring deliberative quality'. Another method would be to employ something like the Discourse Quality Index mentioned briefly in chapter one. However, the research question of this thesis is concerned merely with investigating agency revision in high quality deliberation, and therefore this approach is deemed sufficient.

selection process employed by Demos (in conjunction with Criteria) specified a wide mix of individuals from each area. The criteria this was based on: sex, age, ethnicity and social class, are relevant demographic variables. The problem, however, is that no comparison was made with the region (city) that was being selected for. Because of this, it is difficult to argue that the forum truly satisfies the maxim of representative selection on the basis of legitimacy. However, according to the second dimension in which representativeness matters, the argument that the diversity of opinions and backgrounds can have a significant impact on the nature of the discussion, the case study fares much better. Both London and Bradford forums included a diverse set of individuals according to the designated criteria, and as such, the case study clearly performs well on this second element.

Comparison: the case study performs well on randomisation and diversity of participants, although poorly in terms of being representative of the population in question.

#### *Small Group Discussion*

Although both forums selected 20 participants each, once the sessions began London attracted 19, and Bradford 16 individuals respectively. Each forum then split the participants, (randomly at first) into three separate subgroups. Whilst during presentations and question and answer sessions with expert witnesses, the three groups were brought together; discussion always took place amongst these smaller entities. This continued for all four of the sessions, although after week two a decision was made to alter the makeup of each discussion group. This was done on the basis of making sure all participants experienced full deliberation with all members, as well as to improve the quality of deliberation in each subgroup.

Comparison: the case study performs extremely well on this criterion.

#### *Facilitated Discussion*

The facilitators employed throughout the course of the process were all members of staff from the Think Tank Demos, working in various research programmes within the organisation. Before each session, the organisational group (including facilitators) would convene a meeting, where all topics and objectives of the session would be discussed. This meeting also involved feedback from the previous week, where any

problems or best practice was shared across the three facilitators. Each individual was primed on the various criteria that deliberative behaviour should reflect, and were mindful of the various tasks involved in a full deliberation.

Comparison: the case study performs well on this criterion, although professional facilitators working specifically in deliberation would have been preferable.

#### *Pre-Released Material*

Firstly, the structure of the mini public included an overt and quite intensive educational phase in weeks one and two, where participants were asked to focus solely on understanding key concepts, terminology and regulation. This was supported by a regular stream of pre-released material for each week that included both factual and some more controversial pieces of text for individuals to read. Additionally, each week there were a number of expert witnesses available to the groups for questioning, and beyond this, each facilitator had access to a laptop for researching any issues that came up during the discussion.

Comparison: performs extremely well on this criterion, although there were times when participants had not had time to adequately digest and comprehend the material.

#### *Outcome Influence*

The outcome of the project consisted of a Demos report entitled *Private Lives: A People's Inquiry into Personal and Private Information*, which participants were made aware of throughout the process. The report has been used as the basis of a number of roundtable discussions with government ministers. But more importantly, one of the two major stakeholders, the Information Commissioners' Office (ICO), has used the findings in the report (and particularly the recommended courses of action) to inform a number of policy reviews they were involved in.

Comparison: Demos were bound by the findings, but had now power to act on them; whilst the ICO were not bound by them, but did have limited power to act. The case study therefore performs adequately on this criterion.

## 5.6 CONCLUDING REMARKS

---

To summarise this chapter, it is clear that a number of empirical features commonly associated with deliberative mini publics can have a significant impact on the nature and quality of deliberation that they rely upon. By identifying these key variables, and then applying them to my case study, I have established the areas in which the *People's Inquiry into Personal and Private Information* fares well in promoting deliberation, and those areas in which it performs poorly. The overall conclusion, though, is a positive one. The structure of the mini public chosen for the experimental investigation satisfies most, if not all of the empirical criteria of high quality deliberation to an acceptable standard. Because of this, the next section can then be deemed relevant to the normative model of deliberation discussed in chapter one.

## — CHAPTER 6 —

# RESULTS AND ANALYSIS: DOES DELIBERATION TRIGGER AGENCY REVISION?

### 6.1 INTRODUCTION

---

In chapters two and three of this thesis, I provided an empirically viable hypothesis regarding the impact that deliberation has on the individuals involved in it – namely that it has the potential to alter their agency, and induce them to team reason. In chapters four and five, I then outlined both an experimental approach to test the extent to which this was true, as well as describing the particular case study of deliberation upon which the study was based. I turn now to the results of these experiments.

My concern in this chapter, then, is three fold:

- i. To provide a description of the results of the experimental study using the deliberative mini public.
- ii. To discuss and explain these results on a conceptual level with reference to rational choice and social psychological theory.
- iii. To relate this conceptual explication to the theory of deliberation at the micro level.

To do this effectively, it is therefore relevant to structure the first section of this chapter according to the types of games that were played. Games one to three deal with individuals playing with an anonymous member of the same mini-public; game four deals with individuals playing with a member of the general public; and game five with a member of an alternative specified mini public. As I discussed in chapter four, each of the three types of game is distinct in that whilst it measures the same *type* of deliberative shift, it does so in relation to a very different ‘we’. Because of this, these three sets of results need to be treated as distinct, although the conclusions and explanations derived from them will be linked.

## 6.2 STATISTICAL METHODOLOGY

---

The structure of the results section, then, is quite simple. Each game will first be subjected to simple descriptive analysis, to identify whether any visible changes in strategy choices have occurred over the time period that covers the four deliberative sessions. If a pattern or relationship is identified, the second stage will be to perform a more detailed analysis to ascertain both the extent and statistical significance of this relationship.

Because of the particular characteristics of the data though, these two endeavours are slightly more complex than a simple regression or hypothesis test. More specifically, the binary/categorical nature of the data (in so much as individuals were asked to state only either selfish or team reason strategy choices) means a number of the important assumptions made in standard regression models are no longer met. In particular, it is not possible to assume normality in terms of the distribution of the data – in this case, the binomial structure means neither the mean nor the variance are independent. Logistic, or logit regression, conversely, deals with proportions and probabilities as opposed to continuous information, and transforms the scale of a probability or proportion to a plus or minus infinity.

The logit link function has the following form, with the term in the square brackets equivalent to the odds of any specific event occurring:

$$\text{Logit}(P) = \text{Log} \left[ \frac{P}{1-P} \right]$$

The linear logit model applicable for these results can therefore be expressed by the following equation:

$$\text{Log} \left[ \frac{P_i}{1-P_i} \right] = \alpha + \beta_1 \text{time}$$

In this case  $P_i$  is the probability that an individual will choose the strategy choice ‘team reason’,  $\alpha$  is a constant, and most crucially  $\beta_1$  the likelihood that the given time period

has an impact on determining the strategy choice of the individual. And for non-linear logit regression, the following equation is used:

$$\text{Log}\left[\frac{P_i}{1-P_i}\right] = \alpha + \beta_1 + \beta_2 \text{time}$$

With the appropriate statistical approach now in place, the next step is to consider two issues that are relevant to the way this model is applied to the data. The initial question concerns the nature of the results. Firstly, if time is taken as a proxy for the deliberative event as a single entity, then one simple way of looking at the data might be to imagine the process as a basic pre-test/post-test mechanism. What is therefore important in this conception is simply the overall change (if any) of the strategy choices employed by the individuals. But because of the way the mini public was structured, and specifically that there were four separate sessions corresponding largely to four important phases within the deliberation, it is also possible to look within the data itself. Readings were taken at the following five different time points:

- $t_0$  – baseline measurement (before the first session commenced)
- $t_1$  – after session one (information and learning phase)
- $t_2$  – after session two (questioning and clarification phase)
- $t_3$  – after session three (discussion and debating phase)
- $t_4$  – after session four (discussion, debating and decision-making phase)

The usefulness of this additional layer of investigation is that it allows for a slightly more nuanced analysis to take place, where interesting and important patterns contained within the data can be uncovered. Moreover, when it comes to providing a theoretical justification for any overall relationship, more credibility can be attached to the relevant conclusions if they also support any smaller constituent phenomena. But what should be apparent, of course, is that the deliberative process is cumulative. This means that whilst it becomes possible to test the impact of the deliberative process at, for example,  $t_3$ , the reading does not give the result of that specific session's effect on the individuals in question. Rather, it serves to compare the total effect of all sessions up to that point, in comparison with the baseline reading at  $t_0$ .

## 6.3 RESULTS AND STATISTICAL ANALYSIS

### 6.3.1 GAMES ONE TO THREE

The tabular form for game one is given by the following, with the individual playing the game denoted by individual  $i$ .

		Individual $j$	
		A	B
Individual $i$	A	5 , 5	10 , 4
	B	3 , 10	7 , 7

Figure 4.6 restated: Game Theoretic Encounter One

The results for game one are as follows:

Time Period	Selfish Strategy	Team Reason Strategy
Pre Deliberation ( $t_0$ )	60%	40%
Post Deliberation 1 ( $t_1$ )	54.3%	45.7%
Post Deliberation 2 ( $t_2$ )	42.4%	57.6%
Post Deliberation 3 ( $t_3$ )	48.6%	51.4%
Post Deliberation 4 ( $t_4$ )	48.6%	51.4%

The tabular form for game two is given by the following, with the individual in question denoted by individual  $i$ :

		Individual $j$	
		A	B
Individual $i$	A	7 , 7	2 , 11
	B	10 , 2	4 , 4

Figure 4.7 restated: Game Theoretic Encounter Two

The results for game two are as follows:



Time Period	Selfish Strategy	Team Reason Strategy
Pre Deliberation ( $t_0$ )	60%	40%
Post Deliberation 1 ( $t_1$ )	57.1%	42.9%
Post Deliberation 2 ( $t_2$ )	39.4%	60.6%
Post Deliberation 3 ( $t_3$ )	37.1%	62.9%
Post Deliberation 4 ( $t_4$ )	37.1%	62.9%

The tabular form for game three is given by the following, with the individual in question denoted by individual  $i$ :

		Individual $j$	
		A	B
Individual $i$	A	10 , 10	20 , 2
	B	2 , 16	15 , 15

Figure 4.8 restated: Game Theoretic Encounter Three

The results for game three are as follows:

Time Period	Selfish Strategy	Team Reason Strategy
Pre Deliberation ( $t_0$ )	65.7%	34.3%
Post Deliberation 1 ( $t_1$ )	42.9%	57.1%
Post Deliberation 2 ( $t_2$ )	48.5%	51.5%
Post Deliberation 3 ( $t_3$ )	51.4%	48.6%
Post Deliberation 4 ( $t_4$ )	54.3%	45.7%

Descriptive analysis of the results from games one to three indeed appears to show a change of strategy choices over the course of the four deliberative sessions. It is clear that from  $t_0$  to  $t_4$  in all three games that there has been an overall increase in the percentage of individuals employing the team reason strategy choice. In game one, for instance, there is an increase of 11.4%, game two sees a larger increase of 22.9%, and game three an increase of 11.4%. On the surface, all three games appear to exhibit a relatively modest shift when the percentages are converted into *actual* players; 10%

represents only four individuals, for example. More interesting though, is the shape of this increase.

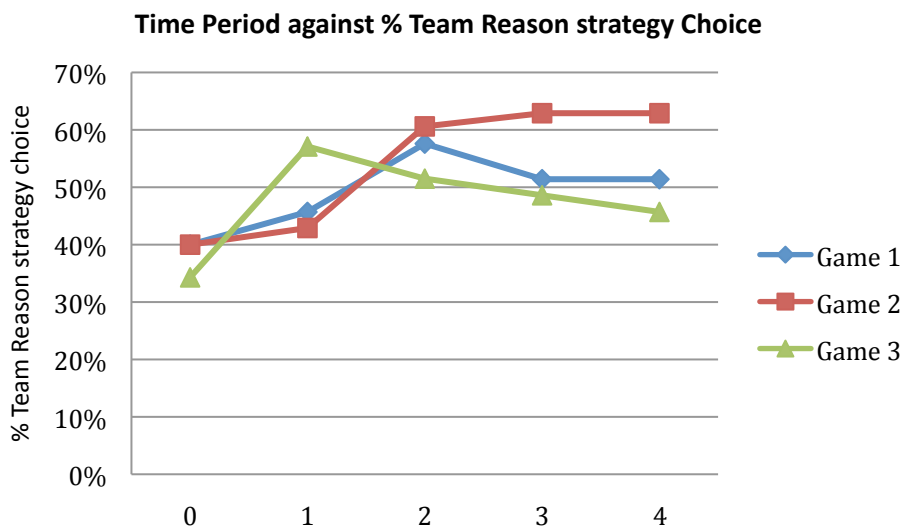


Figure 6.1: Team Reason Strategy Choices in Games One to Three

When displayed on the same axis, it is clear that there is a strong similarity between games one and two. They both start with the same percentage of individuals utilising team reasoning as a strategy; and they both see a dramatic increase at time period  $t_2$ , which is then followed by either by a subsequent reduction or stabilising of this percentage. And although not immediately so obvious, game three in fact follows an almost identical pattern, with the trigger for a large increase occurring at  $t_1$ . Descriptively, then, it appears that all three games suggest an inverse ‘U’ shape relationship exists between time (deliberation) and the percentage of individuals choosing the team reason strategy. The next step, then, is to investigate the extent of this relationship, and equally imperatively, the degree to which these changes are statistically significant.

Running a logistic regression in SPSS for all three games, and treating time as a factor, the following results are obtained:

Table 6.4: Logistic Regression Parameter Estimates for Game One							
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.057	.3382	-.606	.720	.029	1	.866
[Time=1]	-.463	.4294	-1.304	.379	1.161	1	.281
[Time=2]	-.114	.4426	-.982	.753	.067	1	.796
[Time=3]	.248	.4332	-.601	1.097	.328	1	.567
[Time=4]	5.315E-16	.2802	-.549	.549	.000	1	1.000
(Scale)	1						
Dependent Variable: Outcome for game 1							
Model: (Intercept), Time							

Table 6.5: Logistic Regression Parameter Estimates for Game Two							
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.526	.3498	-.160	1.212	2.262	1	.133
[Time=1]	-.932	.5421	-1.994	.131	2.953	1	.086
[Time=2]	-.814	.4719	-1.739	.111	2.973	1	.085
[Time=3]	-.095	.4227	-.924	.733	.051	1	.822
[Time=4]	-4.515E-16	.4579	-.897	.897	.000	1	1.000
(Scale)	1						
Dependent Variable: Outcome for game 2							
Model: (Intercept), Time							

Table 6.6: Logistic Regression Parameter Estimates for Game Three							
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.172	.3393	-.837	.493	.257	1	.613
[Time=1]	-.479	.4105	-1.283	.326	1.360	1	.244
[Time=2]	.577	.4725	-.349	1.503	1.493	1	.222
[Time=3]	.232	.3982	-.548	1.013	.341	1	.559
[Time=4]	.115	.2559	-.387	.616	.201	1	.654
(Scale)	1						
Dependent Variable: Outcome for game 3							
Model: (Intercept), Time							

The  $\beta$ -values calculated for each time period indeed match up nicely to the picture outlined in the simple descriptive analysis, but the truly interesting result from the test is that the significance levels are all extremely low. Statistically then, this can be interpreted in two significant ways. Firstly, the sample size for the investigation, whilst large when considered against the typical number of individuals involved in two citizens' juries, is in fact relatively small when trying to investigate shifts of this magnitude. In essence, it might therefore be true that the pattern *is* reflective of reality, but the sample size is simply too small to make this claim generalisable. The second plausible interpretation, of course, is that the visible inverse 'U' shape relationship witnessed between time and the percentage of individuals choosing to team reason appears merely due to chance – there is no *real* pattern in the data. Both these possible explanations will be considered and discussed in section 6.4 of this chapter.

### 6.3.2 GAME FOUR

The tabular form for game four is given by the following, with the individual in question denoted by individual  $i$ , and individual  $j$  as a randomly selected member of the general public:

		Individual $j$	
		A	B
Individual $i$	A	11 , 11	2 , 16
	B	17 , 3	4 , 4

Figure 4.9 restated: Game Theoretic Encounter Four

The results for game four are as follows:

Time Period	Selfish Strategy	Team Reason Strategy
Pre Deliberation ( $t_0$ )	65.7%	34.3%
Post Deliberation 1 ( $t_1$ )	42.9%	57.1%
Post Deliberation 2 ( $t_2$ )	57.6%	42.4%
Post Deliberation 3 ( $t_3$ )	51.4%	48.6%
Post Deliberation 4 ( $t_4$ )	54.3%	45.7%

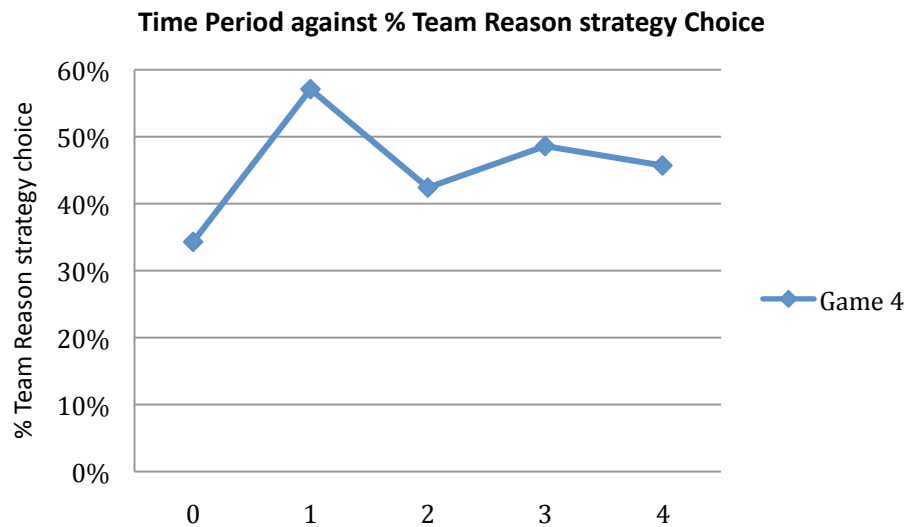


Figure 6.2: Team Reason Strategy Choices in Game Four

Running a logistic regression we get the following results:

Table 6.8: Logistic Regression Parameter Estimates for Game Four							
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.288	.3416	-.382	.957	.709	1	.400
[Time=1]	-.575	.4089	-1.377	.226	1.980	1	.159
[Time=2]	.118	.3904	-.647	.883	.091	1	.763
[Time=3]	-.105	.3459	-.783	.573	.093	1	.761
[Time=4]	2.335E-16	2.4463E-8	-4.795E-8	4.795E-8	.000	1	1.000
(Scale)	1						

Dependent Variable: Outcome for game 4  
Model: (Intercept), Time

As with games one to three, we see an inverse ‘U’ shape relationship appearing between the two factors. But again, whilst this claim is supported by the various odds ratios at each time point (except for  $t_2$  which appears as a slight anomaly), it is once more undermined by particularly low significance levels – leading to the same possible avenues to explore.

### 6.3.3 GAME FIVE

Finally then, the tabular form for game five is given by the following, with the individual in question denoted by individual  $i$ , and individual  $j$  as a randomly assigned member of the corresponding alternative deliberative forum:

		Individual $j$	
		A	B
Individual $i$	A	4 , 4	20 , 2
	B	3 , 20	13 , 13

Figure 4.6: Game Theoretic Encounter Five

The results for game five are as follows:

Table 6.9: Experimental Results for Game Five

Time Period	Selfish Strategy	Team Reason Strategy
Pre Deliberation ( $t_0$ )	62.9%	37.1%
Post Deliberation 1 ( $t_1$ )	57.1%	42.9%
Post Deliberation 2 ( $t_2$ )	63.6%	36.4%
Post Deliberation 3 ( $t_3$ )	51.4%	48.6%
Post Deliberation 4 ( $t_4$ )	51.4%	48.6%

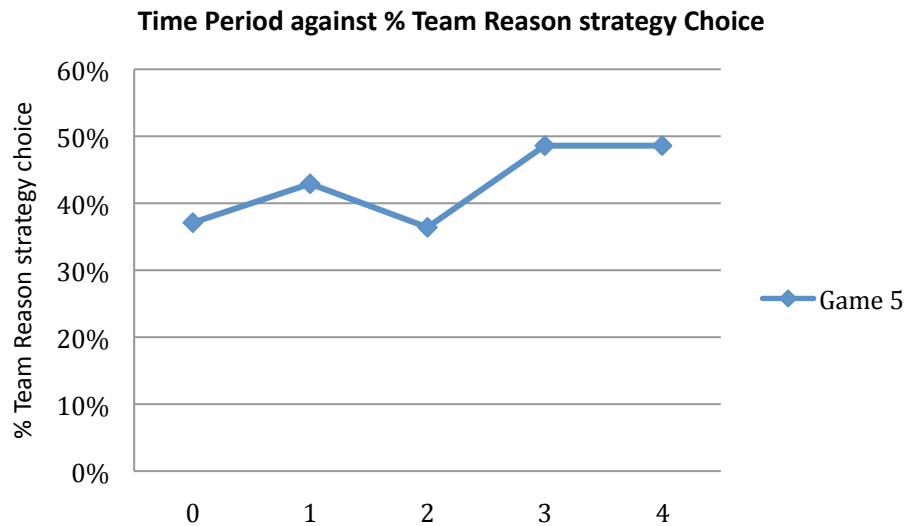


Figure 6.3: Team Reason Strategy Choices in Game Five

Running a logistic regression the following results are obtained:

Table 6.10: Logistic Parameter Estimates for Game Five

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.057	.3382	-.720	.606	.029	1	.866
[Time=1]	-.469	.4354	-1.322	.384	1.160	1	.281
[Time=2]	-.231	.3632	-.942	.481	.403	1	.526
[Time=3]	-.502	.3304	-1.150	.145	2.313	1	.128
[Time=4]	1.200E-16	.	.	.	.	1	.
(Scale)	1						

Dependent Variable: Outcome for game 5

Model: (Intercept), Time

This time, the results are slightly different. Game five does not seem to exhibit the same inverse ‘U’ shape result that the preceding four games showed. Indeed, whilst there is an increase in the percentage of individuals choosing team reason as a strategy overall (again with a slight anomaly at  $t_2$ ), the familiar drop off at  $t_3$  or  $t_4$  does not occur. In other words, the amount of individuals displaying the agency shift appears to stabilise after the immediate increase.

## 6.4 FURTHER ANALYSIS

---

Now that the relevant patterns and statistical significances have been teased out from the data, the next stage is to compare these results with the theoretical approach offered earlier in the thesis. More specifically, how can these results be explained by the rational choice and social psychological approaches discussed in chapters two and three? And from this, how does this then relate to the theory of deliberation outlined in the latter parts of chapter one. To do this, it is imperative to therefore separate this section into two distinct sub-sections. The first assumes that the relationships identified in the data *are* meaningful, but statistically insignificant on the basis of sample size. I will refer to this as the ‘robust’ interpretation of the data during the rest of the thesis. The second proceeds from the assumption that the relationships identified in the results, due to the low levels of statistical significance, are simply the result of chance. This will be labelled the ‘non-robust’ interpretation of the data.

### 6.4.1 THE ROBUST INTERPRETATION

In chapter two then, by drawing on the rational choice approach to decision-making, I sketched a model of deliberative revision that identified three levels at which change might take place. The third layer, that of agency, was then further discussed in chapter three, particularly with respect to the social psychological concepts of identity and social categorisation theory. These arguments can be summarised more simply in the following schemata of propositions:



- Agency (or how people reason) is a clearly defined level at which deliberative revision can take place, and that ‘team reasoning’ is one possible manifestation of the effects of deliberation.
- There are a number of possible reasons as to why this particular shift should be considered favourable, and moreover, that a number of models of deliberative democracy implicitly include this within their arguments.
- That social psychology, in particular, offers a useful explanation of the mechanisms at work to cause this revision.

And finally:

- That deliberation, as a very particular form of discussion, has a number of features that suggest it likely to induce such a shift in how people reason.

The results, however, do not seem to match up precisely with these hypotheses. It is not the case that in all games we see all individuals converge on the team reason strategy. Indeed, whilst there is some shift between  $t_0$  and  $t_4$  during all five interactions in favour of team reasoning, the most we ever see is a 22.4% increase in game three – reflecting an actual change in only seven individuals. Moreover, this increase is certainly not uniform across all the different types of game, where the participant is playing with an individual that represents a different ‘we’ for the basis of a social identity. And further still, in games one to four the increase is not sustained throughout the course of the deliberation, where time seems to also trigger an opposing shift towards individual reasoning after a certain point is reached. How, then, can these be explained?

The answer, I suggest, is best elucidated via an argument that runs on a process of backwards induction. Firstly, I have already outlined the results of the experiments that were premised upon the rational choice approach to decision-making. This comprises the first (already completed) step. Next, I want to return to social identity and self-categorisation theories, in order to gain an understanding of the processes that were at work to generate these results. This should provide a conceptual justification for what occurred during the deliberation, and is the second step of the overall explanation. Finally, the third stage involves an exploration of the potential empirical features that were present during the deliberation itself that reconcile with this justification. Taken

together, these three different elements combine to provide a full explanation for the data.

To start, I want to briefly restate the general argument behind social identity and self-categorisation theories that I offered in chapter three. The premise is that identity is a multiple phenomena, and consists of both personal and social aspects that an individual will cognitively switch between, depending on which one is more salient at any given time. The former, which relates largely to idiosyncratic elements of an individual is characterised by particular behaviour, as is the latter, team reasoning being one such manifestation. A number of factors were then identified from the experimental literature as having a robust effect on raising the salience of social identity, thereby logically inducing individuals to team-reason. These included belonging to the same ad hoc/social group, the presence of an out-group, having common preferences, the use of common language, having shared experiences, participating in face-to-face discussion and the more technical concept of interdependence. These factors were then compared with the six key features identified in the behavioural account of deliberation, as well as with the four phases outlined in the structural account.

For reasons of clarity I am going to structure my analysis on the basis of three key questions that arise from the data. The first relates to all five games, the second to games one to four, and the third involves comparative analysis with game five. Mirroring the approach I took in chapter three, in each question I will consider the conceptual (social psychological) perspective first, followed by an attempt to develop the conclusion via re-engaging with deliberative principles.

Q1. What accounts for the immediate increase in the proportion of individuals employing team reasoning in all five games?

Two general hypotheses suggested in the conceptual stages of this investigation appear to be born out in the early stages of the results. Firstly, that deliberation causes agency revision and triggers people to team reason. And secondly, the suggestion made on the basis of the structural account of deliberation, that earlier phases of the deliberative process (particularly the education and information phase) are likely to have an especially pronounced impact on this revision. From a social psychological perspective, this is represented by the salience of a social identity being raised in relation to that of

the salience of a personal identity in individuals that have shifted strategy choices. In games one to three, this is equivalent to saying that more individuals involved in the process began to view the group as a single entity, and further still, that these individuals then began to exhibit behaviour that was reflective of this altered ‘we’ perspective. What is interesting, however, is the degree to which this predicted revision was also apparent in relation to other levels of social identity. In particular, deliberation also triggered the same self-categorisation process when individuals were paired up with both anonymous members of the general public (game four), as well as other anonymous members of another deliberating group (game five). The interpretation of this result is that the deliberation’s impact, at least early on, seems to be at the higher end of the ‘scale’ discussed in chapter four. It has raised the salience of all three social identities – as a member of the deliberating group, as a member of *any* deliberating group, and as a member of the public/society at large. To provide a robust conceptual justification for this increase in games four and five on social-psychological grounds, I need to introduce two features of behaviour, other than team reasoning, that are triggered by social identity:

- i. Stereotyping
- ii. Social Projection

Let me take the issue of stereotyping first. It is roughly described as the process by which a complex entity is depicted merely in terms of its most prominent or salient features/properties. From this base, two further concepts can be derived. The first, self-stereotyping, is defined as the process in which a specific individual begins to define him or herself in this manner. A significant amount of empirical and experimental evidence exists to support this conceptual hypothesis then, especially in the presence of social identity where individuals begin to ‘take on’ the most salient properties of the group in which they are a member (Hogg and Turner 1987; James 1993; Levy 1996; McKillip et al 1977).

In other words, a collection of individuals who perceive themselves to be members of the same social category or group should tend to stereotype themselves in terms of their common attributes. For instance, when describing themselves, they should endorse attributes that are seen as typical of in-group members, and reject those that are seen as typical of out-group members.

(Simon and Hamilton 1994; p. 699)

Some studies have found that self-stereotyping in favour of the group's characteristics is more likely in cases where there is a significant difference in size between the relative in-group and out-group (Simon and Hamilton 1994). In this example, more self-stereotyping occurred in the smaller group. Others have looked at the impact of expectations in relationships on the phenomenon (Sinclair et al 2006), whilst there is a significant body of work investigating the impact that it has on issues of gender and ethnicity (see Sesko and Biernat 2010). There is also considerable debate over both the speed at which this process happens, and whether individuals are aware of it (Banaji et al 1993). The second derivation is the practice of stereotyping others (Hamilton and Troler 1986, Hamilton and Sherman 1994). This can occur both on a positive dimension, where individuals view others in terms of favourable categorical information, as well as on a more negative one – often referred to as prejudice (Jussim et al 1995).

Linked to both these phenomena is the concept of social projection, which is defined as the process by which 'people come to believe that others are similar to them' (Krueger 2007; p. 2). In short, it is based on a combination of propositions that are strongly linked to social identity and self-categorisation theories, as well as that of stereotyping:

- i. In circumstances that prime social identity, individuals are likely to describe both themselves and others in terms of criteria that are common/most salient to a relevant group membership.
- ii. This process is easier in relation to out-groups, because it involves displacement of much less personal information that might run contrary to the stereotypical features (Ames 2004).
- iii. The perception of members of out-groups is therefore much less accurate as it is based on less information (Ryan and Bogart 2001).
- iv. Social projection occurs automatically in cases that require quick judgements, as well as in situations of high cognitive load and time pressure (Epley et al 2004; Krueger and Stanke 2001).

So how do these concepts relate to agency revision in games four and five? The answer lies in the fact that they are both triggered by a combination of the same factors (Crawford et al 2002; Rydell et al 2007; Spencer-Rodgers et al 2007), as well as the fact

that stereotyping and social-projection play reinforcing roles in the self-categorisation process. What I argue has occurred, then, is the following. The early stages of the deliberation saw discussion that performed well against the ideal maxims of being (i) interactive, being based on (ii) equality and (iii) mutual respect and reciprocity, and was (iv) reason-based according to the (v) public principle. This, as I suggested in chapter three, prompted more individuals to cognitively entify the group, and because of this, self-define in terms of the relevant social identity. Having done so, when presented with non-decomposable games in which they were partnered with members of the same group, the team reason strategy was employed.

A further result of this categorisation, though, was that these factors also triggered a process of self-stereotyping, as the individuals began to think of themselves in terms of the normatively favourable criteria that higher quality deliberation involves. Recall that in games four and five the individuals were partnered with co-players with whom none of them had ever had any contact. The only source of information as to who they were playing with was a single line statement informing them they were a member of a particular group: the general public, or the other deliberative forum. On one level then, each individual was therefore playing each game with a member of a very carefully specified 'out-group'. But interestingly, it is also the case that each 'out-group' can simultaneously be labelled as a potential 'in-group', since on a slightly higher level of abstraction they are both potential social identities available to the participants. This duality, combined with the relative lack of information *actually* known about the other person, had the affect to trigger the processes of stereotyping and social projection in each player's mind. Each co-player was cognitively pictured as having some/all the stereotypical features of each group (for example preferences), which in turn were features the individual in question could also self-identify with. This in turn led to a raised salience of the relevant social identities, and resulted in the team reason strategy choice in these two games. In short, the very fact that the participants had never actually met their co-player in both these games meant they relied on stereotyping and social projection to bridge the cognitive gap.

Q2. What accounts for the subsequent decrease in the proportion of individuals employing team reasoning in games one to three, and game four?

Over the course of the entire deliberation, the increase that is apparent in the early stages is not sustained. In games one to three, a significant drop off is witnessed in the second half of the process. This, quite easily, is conceptually justified via a simple turnaround of the answer to question one. Once the individuals participating in the deliberation had been doing so for a specific amount of time, it seems that the raised salience levels of a social identity in relation to both the deliberating group and the public/society were reduced. At the point where personal identity became more prominent than social identity for these individuals, then, their choice of strategy reverted back to individual rather than team reasoning.

Explaining this via a deliberative justification is slightly more complex, and requires re-engaging with some of the issues I identified in chapters one and three. In particular, I need to return to the two approaches that I used to define deliberation: the structural account, which concerns itself with the relevant phases of deliberation; and the behavioural account, which is concerned with the features that typify deliberative communication. As I demonstrated in chapter five, the case study that was used for the experimentation was broadly structured according to the normative framework I developed in chapter one. The early phases of the deliberation were more heavily weighted towards education and information, the middle phases towards debate and discussion of current issues and potential solutions, and the latter phases directed towards the group making an ultimate decision and set of recommendations. In games one to four, the initial increase in agency revision occurred after the first session, with the reduction then taking place during sessions two, three and four.

The first element of the deliberative justification, then, concerns the nature of the deliberative phase that triggered the reversal in strategy choices back to individual reasoning in these games. As I suggested in chapters one and three, the latter stages of any democratic deliberation should focus the discussion on arriving at a decision – and indeed this is exactly what occurred during the case study. From the second session onwards, the group members were encouraged to participate in the deliberation more actively, and indeed towards session three, they were presented with a number of decision-making tasks in preparation for the final stage. The consequence of this requirement (the difference between deliberation and democratic deliberation) is a change in the nature of the behaviour/discussion between individuals. In particular, as I suggested in chapters one and three, the decision-making stage in deliberation might

uncover, and make more salient, the true nature of the disagreement amongst the participants. By this, I mean that once a group is required to make a collective choice that publicly represents the ‘will’ of all members, then it can become more important for individuals to ensure that the final decision is as close to their desired ends as possible. Whilst this clearly involves a relaxation of the public principle, it need not necessarily go as far as inducing individuals to make statements based on self-interest (Mansbridge 2010). The identification of greater than expected difference between deliberating individuals is also a manifestation of the much weaker relaxation of the principle. Recalling once again the definition of ‘public spirited’, where claims are made on the basis of that which ‘the public at large could accept’ (Chambers 2004; p. 390), there is the possibility then, that some members in the group might represent the minority section of the population that simply cannot accept the proposal or reasons behind it. In this instance, whilst the discussion still performs well against deliberative criteria, the gulf between the preferences of members of the group is enough to raise the salience of personal identity in some individuals.

To make this point a little clearer, let me return to the archetypal example of this in the deliberative literature: that of Rawls’ and Gutmann and Thompson’s discussion of the abortion debate. Imagine two individuals partaking in a deliberation. One individual takes the reasonable moral stance that life begins at conception and opposes the legalisation of abortion. The other acknowledges that some individuals will reasonably reject this belief regarding the starting point of human life, and therefore favours the legalisation. In this example, ideal deliberative behaviour, i.e. the existence of a single conception of the common good articulated through public reason, is impossible. Public-spirited reasoning, then, would consist of individuals making claims that large sections of the public might accept. In this example, both positions are indeed indicative of this, and yet they remain diametrically opposed to each other. Deliberation on the topic of abortion, then, might very well lead to the destruction of any social identity between the individuals when a collective decision is required. Relating this to the case study of *Private and Personal Information* (Bradwell 2010), this was especially relevant when the topics of online information and personal health data were considered. Typically, individuals split into two camps during the discussion of these topics. There were those who trusted government, and therefore had no problem with health authorities storing private data in order to improve the NHS. There were also those who took a much more sceptical, but equally reasonable perspective, and pointed

out a number of recent data scandals as reasons to support their opposition to such data collection and storage. When asked to come to a conclusion on such an issue, then, even reasons that were based on the relaxed notion of ‘public spirited’ heightened this policy/preference difference, and worked to reduce the salience of any social identity that had been built up amongst the individuals in the earlier deliberative phases.<sup>54</sup>

Extending this analysis onto game four then. It also appears that once the deliberation had progressed to a point where personal identity was made more salient than that of a social identity, (with respect to the actual deliberating group amongst some individuals); it then had a significant impact on the related processes of stereotyping and social projection. By exposing the increased level of difference that existed, the cognitive image of the co-player in game four that had been constructed in the early stages of the process became subject to challenge and revision. Instead of playing the game with an individual stereotyped as possessing the idealised features of the deliberative principles outlined in chapter one, co-players were again then cognitively pictured to be much more similar to the individuals actually involved in the deliberation. In turn, these individuals had of course begun to exhibit behaviour that was reflective of some of the relaxations of the principles, including the use of ‘public spirited’ arguments rather than ‘public reason’. Just as a drop in the salience of social identity occurred for some individuals in the actual group, the very same effect (largely due to the recognition of difference) was then translated to game four. In other words, in the same way that stereotyping and social projection caused individuals to team reason in game four during the early stages of the process, the reverse was then true as the deliberation got closer to the decision-making phase.

### Q3. Why does game five not exhibit the same decrease?

Assuming a robust interpretation of the data, I now arrive at possibly the most interesting result. Whilst it is evidently true that game four exhibited a decrease in the number of individuals employing team reasoning towards the end of the deliberation, this ‘drop off’ was not evident in game five. Yet if both of the interactions rely on stereotyping and social projection to cause team reasoning, the question becomes: why

---

<sup>54</sup> As I suggested in chapter three, this is a particular result of the topic under consideration, rather than the deliberative process per se. In cases where less disagreement exists, the decision-making phase in deliberation might in fact help promote social identity. This highlights one area for further investigation.



does it occur in one game and not the other? To answer this, as I mentioned briefly in question one, I need to discuss two important distinctions. The first, and most significant, looks at the apparent malleability of stereotypes, and questions whether they are in fact always easily revised. The second builds on this debate, and introduces a number of different aspects of stereotyping (and social projection) that have been found to impact on this level of malleability. These include the debate between an approach founded on automaticity versus one that prioritises ‘thoughtfulness’ (Bodenhausen 1990; Devine 1989; Doosje et al 1995; Kunda and Oleson 1995; Macrae et al 1994) and a debate between descriptive and prescriptive articulations (Gill 2004). Once these issues have been considered, I will then draw on them to suggest a probable explanation for the results of game five.

A popular way of conceptualising the process of stereotyping then, both theoretically and experimentally, has been the claim that it is simply a shortcut for individuals to take in making comprehensive judgements (Fiske 1998). For example, Bodenhausen et al (1994; p. 49) argue that ‘stereotypes can be viewed as judgemental heuristics that are relied upon by social perceivers whenever they lack the ability or the inclination to think more extensively about the unique personal qualities of out-group members’. Whilst Macrae et al (1994; p. 37) take it one step further, and make the normative claim that the process relies only on ‘the execution of rudimentary skills’. In this sense, stereotyping is suggested to occur most often under cases of extreme time pressure (Freund et al 1985, Dijker and Koomen 1996), under conditions of high cognitive load (Gilbert and Hixon 1991), or when individuals are subject to information overloads (Bodenhausen and Lichtenstein 1987). On the basis of this interpretation, it is therefore logical that once an individual does have the time or cognitive ability (etc), then such judgements will be made on more accurate information, and the stereotype will be revised or rejected quite easily. Indeed, taken in conjunction with the mechanism of social projection, this is a core conceptual assumption of the explanation offered for game four.

However, there is also a substantial body of literature that has challenged this presumption of malleability in stereotyping. For example, Kruglanski and Freund (1983) have investigated the notion of ‘epistemic freezing’, when stereotypes become fixed irrespective of new evidence. They demonstrate that additional information, in contradiction to the established image of an individual, can be ignored, and that original

stereotypes can remain robust under certain conditions. Nelson et al (1996), and Blair et al (2001), on this very line of inquiry, have found particular evidence for this under conditions where stereotyping was considered undesirable. In their studies, they looked specifically at the manifestation of prejudice, and whether individuals presented with contradictory individuating information would revise the negative stereotypes they held of other individuals. They both found considerable evidence for an element of resistance, claiming that categorical evidence (stereotyping) can at times be left unchanged in the face of contestation. In other words then, whilst some types of stereotypes can indeed be undercut by more accurate behavioural information (Kunda and Thagard 1996, Fiske 1998), some might be described as being more 'sticky'. The question this generates, of course, is what determines whether a particular stereotype fits in the former, or latter category.

Often then, the argument of malleability has been used in conjunction with that of automaticity. But as a number of theorists and empiricists have demonstrated, the dynamic process of stereotyping is two-fold (Brewer 1988). On this supposition, Wegener and Clark (2006) have integrated these ideas into a model that distinguishes between 'non-thoughtful' (automatic) and 'thoughtful' approaches. The latter, they argue, is indicative of stereotyping which influences 'judgements when social perceivers are engaged in effortful thinking about target-relevant information' (2006; p. 43). Or put more simply still, stereotyping that occurs when individuals *do* have both the cognitive ability and time to make comprehensive judgements, yet still rely on categorical, rather than individuating information. Arguing that both processes lead to slightly different outcomes, they point to a number of studies in the literature on attitude formation to suggest a more considered approach leads ultimately, to more resilient judgements being made (Darley and Gross 1983; Nolan et al 1999). In particular, they base their empirical investigation on the related results that 'more thinking' leads to stronger attitudes, that remain over time (Petty et al 1995), that are more resistant to challenge (Haugtvedt and Petty 1992), and that more powerfully determine future behaviour (Petty and Krosnick 1995). Their experiments provide quite compelling evidence that the very same result translates to the question of stereotyping. Related to this, Smith et al (2005) have found that increased levels of experience can actually have the effect to enhance stereotyping, because familiar objects are often subject to less critical analysis. In this sense, repeated interaction with an individual may not necessarily lead to the revision of a stereotypical judgement held about them by another

person. And in turn, underlying this result Sherman (1996) demonstrated that whilst during low levels of experience stereotypic knowledge was derived from the identification of group exemplars, as interactions increased this changed to a more abstract image stored and retrieved independently of those it was originally based upon. In short then, stereotypes that are more considered, are more likely to remain stable in the face of new information.

The second issue I want to discuss concerns a differential feature of stereotyping that is present most clearly in the work looking at prejudice, especially on the basis of sexual discrimination (Heilman 2001; Rudman and Glick 2001). When a given individual makes a cognitive judgement on another according to categorical, rather than individual information, stereotyping is occurring. But this, on its own, does not tell us whether the process is purely positive, in that it is concerned merely with how things are, or whether some form of normative element is in play. Heilman (2001), for instance, discusses the impact of normative expectations regarding the behaviour of women on their chances of promotion, concluding that stereotyping which involves some form of value judgement is one of the most significant causes preventing women from moving up the 'organisational ladder'. Broadening this division into a wider context, Gill (2004) argues it is best encapsulated by the respective labels of descriptive and prescriptive accounts of stereotyping:

[...] *descriptive stereotypes*, which purport to describe what group members are typically like ("women *are* gentle"), and *prescriptive stereotypes*, which describe the behavioural standards group members must uphold to avoid derision by the perceiver ("women *should be* gentle").

(Gill 2004; p. 619)

After initially discussing the rather more anecdotal evidence provided in the sexual discrimination literature, the results of three different studies are then reported that provide support for three important claims. These can be summarised as the following:

- i. Descriptive stereotypes can be undercut by more accurate and sustained behavioural information.
- ii. Prescriptive stereotypes, on the other hand, can remain in place even in the presence of contradictory behavioural information.

- iii. It is possible for individuating information to simultaneously reverse descriptive stereotypes whilst having no effect on prescriptive stereotypes.

So, how do these issues relate to the fact that there was no inverse ‘U’ shape relationship for game five? The answer, I suggest, emanates from a conceptual assertion that the dual processes of stereotyping and social projection are different in relation to game five than to game four. Drawing upon the discussion above then, two differential features can be attributed to the type of stereotyping (and consequently social projection) at work when individuals were asked to play the non-decomposable games with members of an alternative deliberative forum. The first relies on the prescriptive/descriptive distinction. At the beginning of the deliberation, all individuals involved were briefed on both the topic and the structure of the process. A crucial part of this involved establishing agreement amongst the participants over a set of discursive norms directly related to the principles of ideal deliberative behaviour. Moreover, this was continually reinforced throughout the actual process by the relevant facilitators. The consequence of this, on one level, was the creation of a normative picture of how a ‘good deliberator’ should act, or in other words, a *prescriptive* stereotype. In the early stages where the behaviour of individuals more closely matched this regulative ideal, due to social projection, this manifested itself in an increased number of individuals self-defining in terms of the social identity. And as Gill (2004) demonstrates, even in the face of contradictory individuating information (such as experience in an actual deliberation), prescriptive stereotypes can persist. In other words then, even once the process had reached the point where ‘lower quality’ deliberative behaviour dominated, individuals presented with game five were still cognitively picturing the ideal stereotype of a ‘good deliberator’ as their co-player.

Further to this, the early stages of the deliberation are also important in relation to the debate on automaticity. As Wegener and Clark (2006) argued, when stereotypes are formed on the basis of a more thoughtful or considered approach, they are more likely to remain stable in the face of contestation. After the prescriptive account was outlined at the start of the discussion then, the early phases/sessions of the process were then likely to affirm the prescriptive stereotype in the minds of the individuals. Because this practice of affirmation involved a much more thoughtful process rather than a simple heuristic, it therefore took on the properties more associated with Wegener and Clark’s model. Taking both these arguments together, it is then possible to explain the more

robust nature of the stereotyping and social projection processes involved in game five. Moreover, because the stereotype remained in place, it then explains why social identity, in turn, stayed more salient - and thus why the percentage of team reasoners remained stable.

#### 6.4.2 THE NON-ROBUST INTERPRETATION

I now arrive at the second possible interpretation of the data, where the low values of statistical significance are now assumed to be central to the results. In this sense, it is no longer simply the small sample size involved in the experimentation that is the main driver behind the low levels of significance, but rather, the process itself that was being examined. From this starting point, I would therefore suggest that the effect of deliberation has been multi-directional, raising the salience of social identity in some individuals at some points, and personal identity in other individuals at other points. Because of this, agency revision might indeed have occurred at points in the process, but the levels of magnitude and consistency that are required to generate a robust relationship are not witnessed.

Explaining this conceptually simply requires a reiteration of the arguments presented in answer to Q2, but extended to all five games across the course of the entire deliberation. Briefly, the suggestion was that the process actually served to heighten the recognition of difference between individuals, with this in turn leading to lower quality deliberative behaviour, and lower salience levels of social identity for some individuals. The only difference, under a non-robust interpretation of the data, is the assertion that this must have occurred immediately, beginning in the very first deliberative phase. Whilst some aspects of the deliberation might indeed have had the effect to raise the salience of social identity then, this was merely 'counteracted' by the difference argument. Consequently, no real pattern emerges from the data.

### **6.5 DISCUSSION AND CONCLUDING REMARKS**

---

In this chapter, then, I have attempted to outline the results of the experiment that tested the impact of deliberation on the issue of agency revision. In doing so, I have performed both basic descriptive analysis, as well as then subjecting each set of results to a more robust statistical test (logistic regression). I have then offered two possible explanations of the data on the basis of two different interpretations of the reasons for

low statistical significance. From these various elements of the analysis, three different results can be drawn out. The first is supported by both the robust and non-robust interpretation, and is therefore considered more substantial. This will be referred to in the next chapter as micro result one:

Micro Result One: Deliberation does not necessarily lead to agency revision amongst the individuals involved in the process, particularly in cases where deep disagreement amongst individuals is present.

Relating this back to second-generation deliberation, this potential result raises some specific concerns regarding the claims made by deliberative theorists. One of the main thrusts towards deliberative decision-making, as a distinct method compared with wholly aggregate accounts, is the argument that the revisionary process leads to better outcomes. As I suggested in chapter two, one of these possibilities is *supposed* to be an increase in the number of individuals employing team reasoning (with the associated benefits this in turn provides). This interpretation and result, then, directly contradicts this assertion. If micro result one is indeed corroborated by larger scale empirical investigation, then deliberative theory must respond accordingly and take this into account. Agency revision, or at least in the direction this thesis has investigated, can no longer be used as an argument for the prioritisation of deliberation as a precursor to collective decision-making. If it does not create team-reasoners, then the three suggestions made in chapter two (that it helps to ‘solve’ social dilemmas, that it increases trust between individuals and ultimately promotes community generation) cannot be upheld. To put it quite simply, second generation deliberative democrats must look for other reasons to normatively justify their approach.<sup>55</sup>

The second and third results that I want to highlight make a slightly different argument, although they are supported only by the robust interpretation of the data and therefore considered slightly less substantial. Respectively, these are:

Micro Result Two: Deliberation can trigger a small degree of agency revision, although the effect matches an inverse ‘U’ shape relationship

---

<sup>55</sup> It should be pointed out that this does, of course, raise some empirical questions for further investigation – particularly whether topics characterised by less disagreement deliver the same result. I will discuss this briefly in the final section of the final chapter.

with respect to members of the same deliberative forum and the general public.

Micro Result Three: Even under conditions assumed in micro result two, deliberation can cause *sustained* agency revision in reference to other members of *another* deliberating group.

Again, I now want to relate these two possible results very briefly back to the account of second-generation deliberation I outlined in chapter one. On this interpretation of the data a limited amount of agency revision *is* triggered by the process, and accordingly, the argument for deliberative models of decision-making on the basis of the three reasons I put forward in chapter two remains. However this is not the entire story, as the ‘shape’ of the results also suggests two possible implications for second-generation accounts of deliberation. The first relates to the results from games one to four. The inverse ‘U’ shape relationship that emanates from the data means that deliberation has a more powerful effect much earlier on in the process. It therefore raises the argument that deliberation should be limited to the point where the agency revision effect is largest.

The second implication I want to highlight relates to the fact that game five (where individuals were partnered with a member of another deliberating group) saw a sustained increase with no drop off after a certain time point was reached. In this case, it raises the related question of whether physically deliberating with all members of the group is both necessary and favourable on the agency revision dimension. If, for example, more team reasoners can be created by restricting the group size and by ensuring as many individuals as possible take part in other deliberations, then the unitary conception of deliberative democracy is called into question.

These three results, and the implications they have for deliberative theory, now form the central substantive points in the final concluding chapter of this thesis. The last stage in the project is to apply these results back to the theory of deliberative democracy, and explore any lessons that empirical political science has regarding the recasting of theory in terms of third generation principles. This completes the overall objective that I outlined in the introduction.

# — CHAPTER 7 —

## CONCLUSION: RECASTING DELIBERATION AND DELIBERATIVE DEMOCRACY

### 7.1 INTRODUCTION

---

In the introduction to this thesis I discussed the importance of the application of empirical political science to political theory, specifically in relation to deliberative democracy. After citing a now famous article by James Bohman (1998), I put the case that deliberative models of democracy, having moved beyond the ‘working theory stage’ (Chambers 2003), were now experiencing what Dryzek (2008) has since labelled the ‘empirical turn’.

Situated in this research agenda, in chapter one I then discussed a useful taxonomy for the way deliberative democracy has developed over the past twenty years, before outlining a clear definition of the central feature of these models: that of deliberation. Having established the parameters of the investigation, chapters two and three then deployed the rational choice and social-psychological approaches to decision-making to deconstruct the normative claim regarding one of the transformative powers of deliberation. In doing so, I identified three levels at which revision was possible. The third level, agency, represented a much overlooked and almost entirely ignored line of inquiry. In chapters four and five, I then set out to provide the details of an experimental approach to test whether deliberation did indeed cause this shift, before discussing the results in chapter six. In this final chapter, I now consider how these results interact with the most recent shift in the theory of deliberative democracy. My concerns, then, are as follows:

- i. To provide a discussion of the recent drive towards recasting both deliberation and deliberative democracy in terms of macro, or ‘third-generation’ principles (Elstub 2010).
- ii. To provide an analysis of third-generation models of deliberation, and demonstrate the most significant problem with them in comparison to second-generation, unitary accounts.



- iii. To explore the implications of the experimental results regarding deliberation and agency revision on this transformation in democratic theory.

In fulfilling these three objectives, I will demonstrate that whilst the shift to a macro conception of deliberation (crucial in third-generation deliberative democracy) might rely on an argument of ‘second best’ in terms of solving the scale problem, this is not the case on all dimensions. Because of this, I conclude the thesis with two assertions. Firstly that deliberative democrats should continue to engage with empirical political science to inform theory, and secondly, that they need not be so resistant to the more recent attempts to recast deliberative democratic theory in systemic terms.

## **7.2 THIRD-GENERATION DELIBERATIVE DEMOCRACY**

---

In the first chapter of the thesis, I proceeded on an approach to deliberative models of democracy that was premised upon a distinction between first, second and third generation principles. After outlining the key assumptions that were categorical of primary incarnations (Cohen 1997; Habermas 1987, 1990, 1996a; Rawls 1993, 1997a, 1997b), I then discussed a number of ways in which second-generation accounts (Dryzek 2000; Goodin 2003; and Gutmann and Thompson 2004) have responded to some quite potent challenges. The most crucial of which has been the question of legitimacy (Parkinson 2006a), and in particular, how deliberation can be made feasible and institutionalised within democratic decision-making. It is this question, more than any other, that has framed what Elstub (2010) refers to as third-generation recasting. How can a deliberative democracy be realised in a large, complex society?

Let me briefly restate the problem of scale put forward in Dahl’s ‘back of the envelope’ calculation that I offered earlier then:

If an association were to make one decision a day, allow ten hours a day for discussion, and permit each member just ten minutes – rather extreme assumptions [...] – then the association could not have more than sixty members.

(Dahl 1970; pp. 67-68)

In an attempt to solve this, I discussed a number of different possible solutions. One, mirroring Rawls, looked at restricting the number of decisions over which full

deliberation was suitable (Dryzek 2001). This was dismissed on the grounds that even one decision would be unfeasible on the basis of Dahl's logic. Another discussed the possibility of deliberation 'within', shifting the burden from external discussion to 'internal reflection' (Goodin 2000, 2003, with Niemeyer 2003). Again though, this was then demonstrated as only a partial solution: face-to-face deliberation is still required at some point in the process. A third possibility suggested representation as the answer, with deliberation either only playing a role in the selection process of candidates (Bessette 1994, Gastil 2000), or alternatively, amongst the elected representatives in 'ersatz deliberation' (Goodin 2000; Parkinson 2007). Both these potential solutions seemed to offer some useful possibilities, although did so by partitioning off sections of the democratic system into separate elements. Largely on this basis, I then suggested briefly that a 'third-generation' attempt to recast deliberative democracy in terms of more 'macro' level arguments could also be used to circumvent the issue. It is this more recent reformulation that I now discuss, a shift that Dryzek (2008; p. 1) labels the 'systemic turn' in deliberative democracy.<sup>56</sup>

### 7.2.1 MACRO ACCOUNTS OF DELIBERATION

The most crucial point that macro models of deliberative democracy must confront is the definition of deliberation they make central in their framework. By this, I do not mean the debate regarding the criteria of ideal deliberative behaviour (the behavioural account), and nor do I mean the various stages that a deliberation must be constituted of (the structural account) – although these will become relevant shortly. Rather, I refer to the assumption that deliberative decision-making requires the simultaneous input of all members affected by the relevant issues under consideration. In micro deliberation, which relies on this unitary conceptualisation, this is made possible by an extremely small number of individuals being involved. Macro accounts, on the other hand, claim that the issue of scale is central, and that deliberation must be conceived in a different manner if it can be utilised on a societal level. Four different (although related) theoretical ways of conceptualising deliberation have underpinned the third-generation literature as a mechanism to circumvent the problem of scale, which I now want to discuss. The first two are based only on dividing up the larger deliberating group into smaller subsets, whilst the second two also introduce the idea of a division of

---

<sup>56</sup> I have left the discussion of this reformulation until the final chapter for the very reason that my results speak directly to it. By providing an outline of the approach first, the manner in which an agency revision argument can be deployed is much easier to construct and comprehend.

deliberative labour amongst the deliberators. Once I have done this, I will then discuss how they interact more broadly in terms of deliberative democracy, before then considering how my results interact with this recasting.

### *Parallel Deliberation*

The first conceptualisation, which I term *parallel* deliberation<sup>57</sup>, requires the participation of all individuals in society, but proceeds by splitting the larger group into smaller, more manageable subsets of the population. In doing so, it allows the smaller groups to form their own distinct forums, where higher quality deliberation is therefore made possible. Once individuals have been through *a*, rather than *the* discursive process, they are then required to take part in a common separate aggregation phase, such as voting. The following diagram might thus represent parallel deliberation that involves splitting the population into three different subgroups:

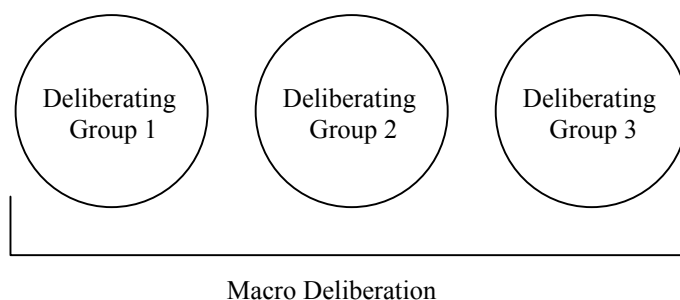


Figure 7.1 Parallel Deliberation

Whilst parallel deliberation therefore involves the full participation of all individuals affected by any decision, it does so by relaxing the requirement of simultaneous contributions in a unitary (or single) forum. In doing so, one aspect of the legitimacy critique is therefore dealt with quite nicely. If those affected by the issue are the source of legitimacy for any collective decision regarding it, then it stands to reason that the more individuals involved in the process, then the more legitimate the outcome. However, by solving this problem, parallel deliberation appears to generate a secondary, related epistemic concern, which I will term the ‘transmission problem’. By constructing lots of different deliberative forums, each comprised of different individuals with different initial preferences and different personal experiences, the suggestion is that the transformative power of the process is likely to be different at

---

<sup>57</sup> Goodin (2000; pp. 87-89) discusses this type of approach in his section on ‘ersatz’ deliberation, although does so from the perspective of using one forum to represent the entire population, rather than each forum doing so.

each site. In turn, this poses a number of questions regarding the nature of the post deliberative outcomes. Has deliberation at one site, because it has taken place without the participation of a particular individual, resulted in a different outcome compared with another? If this is the case, then clearly some forums will be better or worse than others. In particular, those that are attended by a more representative cross section of the population would logically seem to provide for higher quality deliberation. Parallel deliberation thus seems weaker, or less preferred, than a unitary conception on this line of reasoning.

But responding to this briefly, there are two possible counter points that help to mitigate this problem. The first simply draws on the idea of representative selection that is found in the literature on mini publics that I discussed in chapter five. Parallel deliberation could plausibly rely on a formal allocation process of membership, to ensure that every deliberating group involves a representative selection of individuals affected by the decision. Secondly, it is also important not to conflate individuals, with preferences and reasons. Whilst it is possible that every individual might possess a unique set of preferences, values and personal experiences, it is likely that some, if not all, will be duplicated by other members of the forum.<sup>58</sup> Indeed, factor analyses of various real world deliberations has often only uncovered at most five to six possible clusters, with three or four being more common. If this is the case, then whilst parallel deliberation might indeed create some new drawbacks through alleviating the scale problem, these issues do not seem to completely demand the dismissal of the conceptualisation as unworkable.

### *Disjointed Deliberation*

The second approach I want to consider builds upon the idea of splitting large numbers of individuals into smaller sets of deliberating subgroups. In parallel deliberation, the sites or forums are conceptualised as highly distinct and separate. Members of one forum are not required, nor are they welcome, to attend any other. But as I have suggested, this can generate a secondary concern when the quality of the substantive outcome is considered. If a particular perspective is not represented in the discussion, for example, then the final post deliberative preferences arrived at by that particular subgroup may be ‘less’ deliberative on the basis that plurality of inputs is deemed favourable. A formal condition of representativeness is one way of partially

---

<sup>58</sup> I owe this nice counter-point to John Parkinson.

circumventing this critique. An alternative, however, relies on two slight amendments to the approach. The first is an even greater relaxation of the unitary condition. Parallel deliberation goes some way by not requiring individuals to contribute to the same forum, but it is still premised upon the fact that multiple deliberations are to take place at the same time. A further weakening of this condition, allows for the possibility that different forums can take place over the course of a time *period* – allowing some to finish before others even begin.

Built upon this, the second differentiation utilises the idea of a crossover, where membership of the various deliberating forums overlaps with each other. Goodin (2000) labels this type of process as ‘disjointed’ deliberation, and cites Aristotle’s suggestion of deliberating ‘not all in one body, but in turns’ as the original formulation of the approach. Recasting deliberation in these terms again clearly circumvents the scale problem, and does so in way that partially self-mitigates the transmission problem associated with the parallel articulation. By linking deliberating groups with each other, the process allows for the transmission of any unique perspectives and preferences, and goes some way to ensure that each individual gets to deliberate with every other individual – albeit in a rather indirect manner. The simplest representation, then, would be the following:

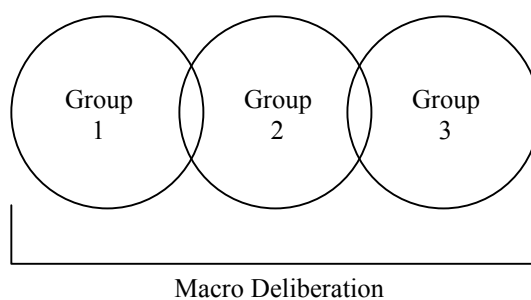


Figure 7.2 Serial Deliberation

However, when disjointed deliberation takes this linear form, the serial nature of the process becomes highly susceptible to the problems of inequality and path dependency.<sup>59</sup> If deliberation is undertaken in this way, those participating in group one will have a disproportionate amount of influence over the macro process in comparison to the members of group three, violating the maxim of equality. This is because whilst individuals involved in deliberating groups towards the end of the time period are

---

<sup>59</sup> This particular example of disjointed deliberation is better termed ‘serial deliberation’.

exposed to the post-deliberative preferences and opinions of individuals from earlier groups, the reverse is not true. A solution to this problem is to allow individuals from different groups to interact with each other throughout the course of the entire process, thereby creating a host of different ‘starting points’ for the macro deliberation. Diagrammatically, this alternative could therefore be depicted by the following:

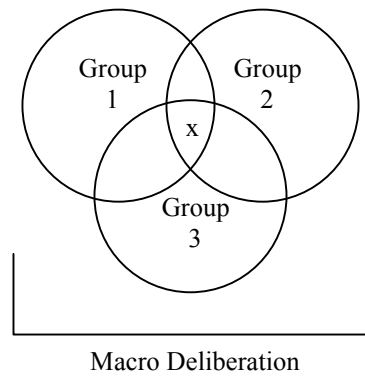


Figure 7.3 Disjointed Deliberation

The most significant difference regarding this non-linear account of disjointed deliberation is that no single group forum represents the start, or finish, of the overall process. There is no definite first or last sub-deliberation. To facilitate this, two foundational assumptions of disjointed deliberation must be maintained. The first is that the different subgroup deliberations must take place over a long enough time scale to allow sufficient overlapping between different forums. And the second is that these respective forums must therefore last long enough themselves to allow individuals to move between the groups. If both these conditions are met, then it is possible to imagine something like a dynamic network of different deliberating groups, each enjoying input from, and input into, other forums. As Goodin himself puts it, ‘there might be a “web of group affiliations” that links (indirectly: perhaps very indirectly) everyone with everyone else in a dialogue that effectively straddles the entire community’ (Goodin 2000; p. 87).

As with the parallel conception though, disjointed deliberation can be critiqued on a number of levels that result in the claim that it is less preferred than the unitary account relied upon in second-generation models of deliberative democracy. Most obviously, the argument levelled at parallel models regarding the transmission of perspectives is only ever partially answered in the disjointed articulation. It cannot, for instance, *guarantee* that all members hear all viewpoints. An indirect transmission does not

function as efficiently as a direct one, and because of this, overlapping membership is clearly a case of the ‘second best’ alternative. From this, the question of how cross membership of different groups actually works also raises some interesting related concerns. Consider the area marked ‘x’ in figure 7.3 above. This represents the collection of individuals who are involved in all three of the group deliberations, and is the basis of this partial solution to the transmission problem. But in turn, this poses the question of *which* individuals should occupy this position? Whilst it seems logical that any basis of membership of this subset should be related to ensuring the transmission of different perspectives, it becomes extremely difficult, impossible even, to formalise this operationally. Any selection process that solves this must start from an impossible point of complete information in order to ‘know’ which perspectives are unique. And to complete the circular nature of this issue; the only real way to guarantee the ‘correct’ or ‘best’ people hold this position is to make the overlap large enough that the scale problem might become an issue once again. In short, whilst disjointed deliberation mirrors the parallel conceptualisation in solving the scale problem, and performs slightly better on the self-generated issue of the transmission problem, it still represents a less preferred articulation than the (small scale) unitary account.

### *Distributed Deliberation*

The third macro account of deliberation that has been utilised in third generation models of deliberative democracy draws upon both the previous two articulations, but takes a slightly different focus. Parallel and disjointed deliberation work by splitting the population into smaller constitutive groups, in order to reduce the number of individuals in each deliberative forum. Distributed deliberation (Goodin 2005) takes this one step further, and is rooted in the assumption that different elements of the deliberative process can be apportioned to these different sections of the population. The simplest way this can be outlined, then, is by way of the structural account of deliberation that I offered in chapter one.<sup>60</sup> Let me restate the four different stages that were identified as essential components in any democratic deliberation:

- Education and information phase
- Identification of solutions phase
- Evaluative criteria phase

---

<sup>60</sup> Parkinson (2006; p. 169) follows a similar logic in his table (the distinction between define, discuss, decide and implement) although with reference to deliberative *democracy*, and thus includes many other forms of political participation and political institutions in his discussion.

▪ Decision-making phase

The first three of these elements are, if you like, the essential aspects of the ‘talk-centric’ part of the process, which the scale problem so applies to. Distributed deliberation therefore works by (a) splitting the population into smaller groups, and (b) by then assigning each group to a particular task. In doing so, the deliberative work is therefore shared out amongst the entire population, with different individuals performing different parts of the macro deliberation. Further to this, it is also possible to then unpack the distributed elements themselves. If, for example, the number of individuals involved in a single phase is still too large, then disjointed deliberation might be employed *within* this stage. Mirroring the previous two discussions, this might be represented by the following:

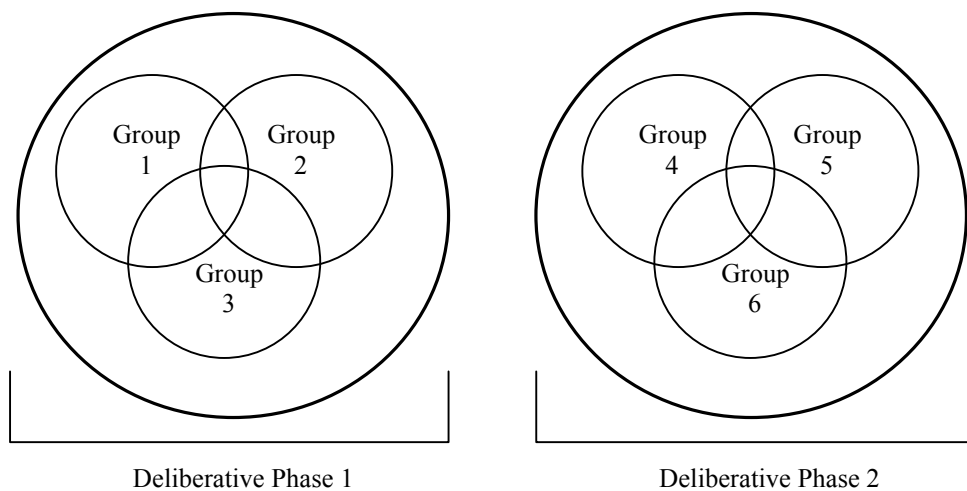


Figure 7.4 ‘Parallel’ Distributed Deliberation

The dilemma for this ‘parallel’ distributed articulation, however, is that the various stages of a deliberation need to be linked together, and in a particular order. For example, the evaluative criteria phase relies heavily on the identification of solutions phase. Without any possible courses of action identified, it is more difficult for individuals to come up with an agreed way of ranking them. And similarly, it would make little sense to come up with the said criteria before a host of possibilities had been first generated. There are two related possibilities that emanate from this. A more structured perspective would argue for a formal transmission mechanism between the various deliberative moments. As I will discuss in the following section regarding macro models of deliberative democracy, this is often the role that formal political institutions play in large complex societies. For example, it might be possible to



imagine the collection of individuals in groups four, five and six in figure 7.4 to simply take the input from groups one, two and three as fixed. If phase one represents the identification of solutions phase, this would mean that *only* those possible courses of action that had been recommended by the earlier deliberating individuals would be considered in the subsequent stages of the macro deliberation.

It turns out, though, that this formalisation merely creates a different version of the transmission problem applicable to parallel and disjointed articulations. Whilst it might be practicably possible to split the stages of deliberation into distinct serial phases, this ignores the fact that many of them are intrinsically linked in *both* directions. Take the relationship between the identification of solutions phase and the establishing of evaluative criteria. It is conceivable that a situation where a particular discussion and decision on the latter might positively influence the outcome regarding the former. What happens say, if the individuals in phase two want to rank possibilities on a particular dimension, yet the deliberation in phase one has paid no attention to this criterion when coming up with different courses of action? The resultant output from the first phase is thus clearly sub optimal from a deliberative standpoint that values plurality of input into the process. Again, a partial solution is to allow for an overlap between the various stages, resulting in an approach that might be labelled ‘disjointed distributed deliberation’. For the third time though, this represents only a second-best solution to the transmission problem when compared with that of a unitary account.

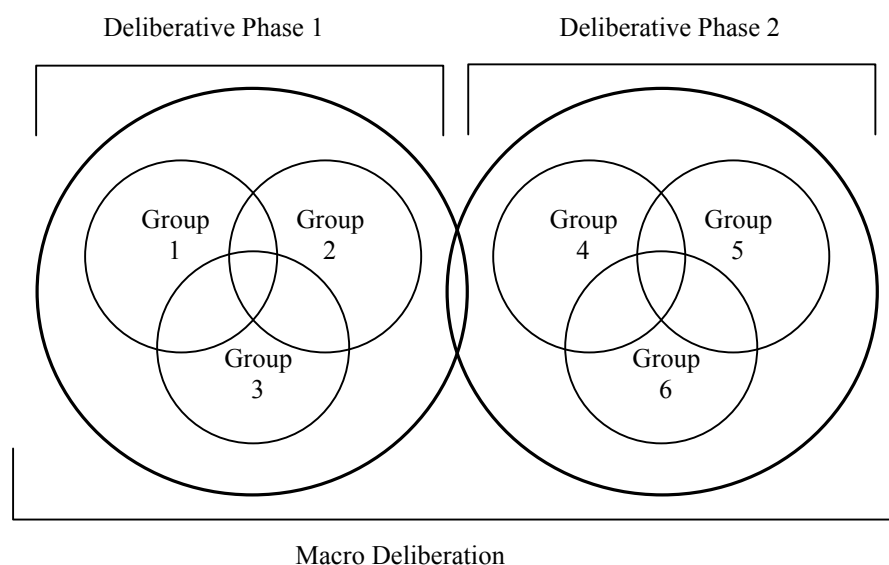


Figure 7.5 ‘Disjointed’ Distributed Deliberation

### *Sequenced Deliberation*

Dividing deliberation up on the basis of the structural account to make it applicable on the macro level is though, only one way in which deliberation might be segmented on the basis of a division of labour. A second starts from the ‘disjointed distributed’ articulation discussed above, but also introduces the behavioural account into consideration. Mansbridge (1999, 2010a) and Goodin (2005) both suggest that the burden for all individuals to exhibit ideal deliberative behaviour for all phases of even a disjointed distributed deliberation is too great for implementation on a macro level. Instead, they construct a model which allows the ‘deliberative quality’ to vary at certain points during the overall process, with particular phases required to perform better according to certain maxims. To make this a little clearer, let me restate the six key features of deliberative behaviour that I identified in chapter one:

- Interactive Communicative Process
- Equality
- Mutual Respect and Reciprocity
- Reason-based Discussion
- The Public Principle
- Decision-focussed

Goodin’s argument is therefore quite simply that not all six criteria need to be completely satisfied by all individuals in all phases. Consider, for example, the education and information phase in comparison with that of the identification of solutions. It is quite logical to expect the latter to involve *more* interactive communication than the former, and similarly, for the former to involve discussion that is non-decision focussed. What matters for sequenced deliberation then, is that the overall process exhibits instances of all six criteria, not that all the individual component phases do. It is on the aggregate level that deliberative behaviour must reflect these six ideals. As Mansbridge (1999; p. 224) puts it: ‘the criterion for good deliberation should be not that every interaction in the system exhibits [these ideal criteria], but that the larger system reflects these goals’.

Out of all four macro interpretations, sequenced deliberation represents the format most removed from the unitary conception offered in second-generation accounts. It includes division by group, division by deliberative phase, and division by deliberative behaviour. But as with the other three formulations, solving the scale issue comes at a

price. Firstly, because sequenced deliberation draws on both distributed and disjointed articulations, it is equally susceptible to the suggestion that it only partially ensures full transmission of all perspectives to all individuals throughout the process. Moreover, because both variants are applicable, the problem is made even more pervasive and entrenched in this approach. For a final time then, overlapping between individuals and deliberative phases provides only a second-best solution in comparison to the unitary perspective.

But secondly, there is also a salient issue regarding the argument for the transformative power of deliberation that must be tackled by proponents of sequenced models. If a particular group of individuals are only required to participate in certain phases, and assuming the allocated phase for an individual is one that performs poorly on deliberative criteria, then clearly some of the positive effects which deliberation is expected to deliver will not obtain. Indeed for those individuals who only participate in the decision-making phase, for instance voting, it is difficult to see how the process is any different from the aggregative model of decision-making which deliberative forms are proposed to replace. For the moment though, I want to put this issue to one side, as I will return to it in section 7.3 when I consider how the results of the agency revision experiments (one such transformative example) relate to these approaches to macro deliberation and deliberative democracy.

Type of Deliberation	Split into groups	Overlap between groups	Division by deliberative phase	Division by deliberative behaviour
Parallel	+	-	-	-
Disjointed	+	+	-	-
Distributed	+	+	+	-
Sequenced	+	+	+	+

Table 7.1 Features of different models of macro deliberation

At this point, it might be useful to briefly summarise where my argument has reached. So far, I have discussed four different conceptualisations of macro deliberation, in particular highlighting the problem of ‘transmission’ that they generate in departing from the unitary account. My next concerns are firstly a discussion of how deliberative models of democracy might be conceptualised using third-generation principles, and

secondly, how these different macro interpretations of deliberation have been incorporated into them.

### 7.2.2 MACRO ACCOUNTS OF DELIBERATIVE DEMOCRACY

As I have asserted numerous times already, the shift from focussing on micro to macro models of deliberative democracy represents the most recent development in the field of study. As such, the relative paucity of literature in the third-generation tradition reflects both its short history, as well as the fact that it symbolises what some could consider a retreat from the ideal deliberative project. Arguing for a recasting of deliberative democracy that draws on a macro interpretation of deliberation might appear as partially surrendering the significant body of work that has sought to establish and entrench the deliberative turn. This is something that I will address much further in the following section (7.3), but for now, I want to provide a brief description of some of the more cogent third-generation accounts.

Let me start by briefly restating the general claim(s) regarding macro models of deliberative democracy. A deliberative democracy understood through a micro lens, might be envisaged as a unitary society of individuals who are required to discuss, and then vote on a particular issue. Citizens' juries, consensus conferences, deliberative polls etc, are all direct manifestations of the guiding principles behind this approach. They aim to bring together all the individuals affected by an issue, and then use deliberation and aggregation as a way of securing a legitimate collective decision. Macro accounts, on the other hand, take the size and complexity of large societies as the starting point for their model, and construct a normative framework for political decision-making that draws on a host of different institutions. In the same way that macro deliberation works by separating either the individuals involved or the deliberative process itself into smaller elements, third-generation models of deliberative democracy work by splitting the democratic system up into various smaller units. In this sense, they aim to describe a system of governance and decision-making that is both more applicable, and similarly more closely reflects, contemporary democratic political systems.

Discussions regarding macro models of deliberative democracy have largely centred around two related issues. The first is *which* institutions should be included in any formulation, whilst the second asks how these various institutions should be linked with

each other. But before I continue with this discussion, I want to draw upon a distinction made by Dryzek (2008), who examines how a deliberative democracy might be envisaged on the largest scale of all – that of a transnational level. In it, he argues for a differentiation between three different planes of formality in describing a macro democratic process.

The first, which relies heavily on an idea also presented in Hendriks' (2009), is the idea of a deliberative democratic 'soup'. According to this interpretation, a macro model of deliberative democracy should include a host of different institutions that satisfy both deliberative and democratic functions, but do so in an entirely unstructured and non-systematic manner. For example then, a 'soup' might include the use of mini publics that perform well on deliberative criteria, and it might also include an election process for a legislature that involves all citizens in an equal manner. What it does not include, however, is any normative, formalised link between these two institutions. A deliberative democratic soup pays no attention to the order, or how the mini publics might be linked to the election process: all that matters is that both institutions merely exist. The second interpretation, which he terms a deliberative 'society', is slightly more structured than a soup, because it 'has norms and discourses that regulate the activities of and interactions of all the relevant composite members' (Dryzek 2008; p. 5). In this sense, a deliberative society, whilst it does not suggest that institutions should be structured according to a formalised normative framework, does argue that the individuals involved possess certain values and norms. These values and norms are expected to guide both the behaviour of individuals, as well as provide *some* basis for how the institutions are related with each other in the overall process. The crucial point to draw out though, is that because the values and norms are socially constructed, they are free to change at any point. To put it more simply, a deliberative society relies on endogenously determined criteria to guide the deliberative democratic decision-making process.

The final interpretation, and the one which has proven the single most popular in democratic theory, is the concept of a deliberative democratic 'system' (DDS). The term, first utilised by Mansbridge (1999) in an article discussing the importance of 'everyday talk', represents the most structured way of conceptualising a deliberative democracy on the macro scale. A system, as with a 'soup', includes a number of different institutions that perform various different deliberative and democratic

functions alike. It also relies on certain values and norms to help shape the process, mirroring the idea of a deliberative democratic society. Where it differs, however, is the fact that some of these norms are now expected to provide the explicit structure, and required rationale, of the decision-making process. More specifically, a deliberative democratic system should function in a manner as close as possible to the ideal requirements established in second-generation, unitary accounts. Numerous deliberative democrats have since offered different accounts of deliberative democratic systems, including Mansbridge (1999), Ackerman and Fishkin (2002, 2005), Goodin (2005), Hendriks (2006) and Parkinson (2006a).<sup>61</sup> Whilst differing in their respective focus and approach, all share five common characteristics built into their models. These include a public space, an empowered space, a transmission mechanism between them, the accountability of the latter to the former, and the opportunity for individuals to deliberate over the make up of the system (Dryzek 2008; pp.8-9). Very briefly, let me discuss the three most relevant to my overall argument.<sup>62</sup>

### *Public Space*

The idea of public space, in which individuals are free to associate and communicate with each other, is akin to Habermas' (1996b) articulation of the 'informal public sphere' in his two-track model of politics. Although modelled as having no formal power to actually make collective decisions, public space is argued to discharge a number of other essential democratic functions. As Habermas so succinctly puts it, then:

[...] new problem situations can be perceived more sensitively, discourses aimed at achieving self-understanding can be conducted more widely and expressively, collective identities and need interpretations can be articulated with fewer compulsions than is the case in procedurally regulated public spheres.

(Habermas 1996b; p. 308)

Relating this to the literature on deliberative systems, Mansbridge (1999), for example, focuses on the role that 'everyday talk' plays in satisfying this criterion. Goodin (2005), conversely, describes a deliberative system specifically in relation to a liberal representative democracy, and relies solely on the free election campaign to fulfil this

---

<sup>61</sup> The phrase can also be found in Thompson (2008), Bohman (2009) and Mansbridge (2010b), which gives an indication of the general direction of the field of study.

<sup>62</sup> The first three represent the substantive features that allow me to then consider the 'type of macro deliberation' which they rely upon, in the following section.

element. For Parkinson (2006a) and Hendriks (2006) on the other hand, who sketch a slightly more institution-based articulation, this might also include things like participation in activist networks, political protests, or the media. Finally, Ackerman and Fishkin (2002) offer perhaps the most radical account, when they argue for all citizens to attend separate forums to participate in structured deliberations to guide subsequent political action. Different formulations might include different institutions then; but they all have at least one that performs this function. Importantly, within a deliberative system the fact that this space is open and available to all citizens has significant impact on both the nature of participation found within it (which I will consider in the next section), and the role it plays in the wider process. Habermas, Ackerman and Fishkin, Mansbridge, Goodin, Hendriks and Parkinson all suggest in various guises, for instance, that the institutions found within ‘public space’ are especially effective at securing the ‘democratic’ foundation for the deliberative democratic system.

### *Empowered Space*

The notion of empowered space again draws upon Habermas’ work (1996b), but this time is representative of the arena he referred to as the ‘formal political process’. This second element of the two-track model relates directly to more conventional political institutions that are required to make the binding collective decisions. In Goodin (2005), Parkinson (2006a), and Hendriks (2006), this remains the prerogative of an elected representative assembly. But it is possible to also consider other ‘less’ democratic institutions that might be asked to perform this function. For example in constitutional democracies, it is often the role of the appointed judiciary to provide the final judgement as to whether a collective decision is legitimate. Something like a Supreme Court, then, might fulfil this requirement. Ackerman and Fishkin’s (2002, 2005) focus on the Presidential election is perhaps the least ‘democratic’ example in action, as it reflects a single individual who makes up the arena.

The privileged nature of this forum, compared with that of public space, means it has subsequent different resultant obligations placed upon it. For instance, the fact that it is composed of a much smaller number of individuals means it is much more likely to be able to secure higher quality (even unitary) deliberation. Because of this, most accounts (notably not Ackerman and Fishkin’s, though) tend to suggest this institution as the element best (although not uniquely) equipped to provide the main ‘deliberative’ thrust

of the deliberative democratic system. If public space is to provide most of the ‘democraticness’ then, empowered space is argued to add the ‘deliberativeness’.

### *Transmission Process*

In establishing two distinct elements of a larger deliberative democratic system, the obvious next issue to consider is how they are linked. This is essential for the process to work as a ‘system’ rather than a ‘deliberative democratic soup’. But more importantly, political legitimacy is dependent on the transmission process between these two respective segments of the larger process. As Cohen (1999; p. 409) argues: ‘the two-track model indicates how (communicative power) might flow from citizens, reasoning in a dispersed network, through a deliberative legislature, to administration’.

As I suggested in section 7.2.1 regarding sequenced deliberation, deliberative democratic systems rely on certain types of institutions to provide this service. Ackerman and Fishkins’s (2002 2005) model offers perhaps the simplest example of this in action, where preferences constructed in the public space are fed into the empowered space via the casting of electoral ballots. Taking a slightly different (and more deliberative) approach, Hendriks (2006; p.500) refers to these as ‘mixed discursive spheres’, whilst Parkinson (2006a; p. 166) prefers the term ‘middle democracy’. Both, however, appeal to rather similar instances of the process in action. They suggest examples including mini publics (citizens’ juries, consensus conferences etc), facilitated town hall meetings, public seminars and even the media. The key point to take, however, is that Hendriks and Parkinson favour a mixed approach to this aspect of the system.

### 7.2.3 DELIBERATION IN DELIBERATIVE DEMOCRATIC SYSTEMS

Dryzek’s (2008) classification of the features that deliberative democratic systems incorporate offers a useful analytical framework for describing these models. What it omits to consider, however, is the type(s) of deliberation they each rely upon. In this section, I want to address this shortfall, and discuss which of the various conceptions of macro deliberation detailed above are present in deliberative democratic systems.

The overall picture of the accounts offered by Mansbridge (1999), Ackerman and Fishkin (2002, 2005), Goodin (2005), Hendriks (2006) and Parkinson (2006a), then, is one of a sequenced deliberation. But within this, it is also possible to identify the



presence of other formulations. Consider how deliberative democratic systems work. They begin by segmenting the deliberative democratic process into different arenas, which in turn fulfil different phases of the structural account of deliberation. From this, and because certain arenas/phases clearly lend themselves to performing better according to different ideal behavioural criteria, a subsequent division on this dimension is also undertaken. Moreover, not all individuals are required to participate in all of the different institutions or arenas, although a degree of overlap is present. From a holistic perspective, a deliberative democratic system therefore ticks all four of the boxes that define a sequenced deliberation.

But this is not the entire story. Looking at the two distinct arenas as well as that of the transmission process, it is clear that different accounts of deliberation dominate in each of them. In the formal empowered space, the relatively small number of individuals expected (or indeed allowed) to participate facilitates the use of the ‘gold standard’ – it is able to proceed on the basis of unitary deliberation. Conversely, the public space, because of the requirement for it to act as the main democratic element in the system, involves far more individuals. As a consequence, it must rely on either disjointed or parallel deliberation to ensure all individuals are able to participate in a deliberative fashion. And finally, within Parkinson (2006a) and Hendrik’s (2006) descriptions of ‘middle democracy’ or ‘mixed discursive spheres’, it is a mixture of all types of deliberation that predominates. In this element, the result of the large-scale disjointed process is taken into numerous different deliberative forums before it is finally fed into the unitary empowered space.

Deliberative democratic systems, then, are examples of a sequenced deliberative process, which in turn is constituted by unitary, parallel and disjointed formulations of macro deliberation. Most crucially, because of the associated ‘transmission problem’ generated by the parallel and disjointed articulations, the argument that a deliberative democratic system represents a less preferred account of deliberative democracy than that of a wholly unitary conception can of course be levelled. The problem, to put it quite simply, is that the only guaranteed way to ensure the contestation of all preferences and perspectives is to allow all individuals to deliberate with each other. Relaxing this maxim, as I argued in section 7.2.1, solves the scale problem at the cost of this requirement. To sum up this argument in Parkinson’s (2006a) words then:

[...] while we can imagine a deliberative system that is more legitimate than current arrangements, no one event can ever be fully legitimate and at the same time strictly deliberative, because not all the elements of legitimacy, democracy and deliberation can be present in one process.

(Parkinson 2006a; pp. 174-175)

### **7.3 AGENCY REVISION AND DELIBERATIVE DEMOCRATIC SYSTEMS**

---

Again let me briefly summarise where my argument has reached. So far in this concluding chapter, I have presented an outline and analysis of the different accounts of macro deliberation and approaches to macro deliberative democracy. I have then suggested that because the latter so relies upon the former, that the deliberative problem of ‘transmission’ (which is generated in an attempt to solve the issue of scale) can therefore be used to critique deliberative democratic systems on the same basis. For this reason, the shift from second to third generation-accounts of deliberative democracy might be seen as one of compromise. As Parkinson argues, ‘the ideal remains the full involvement of every citizen in every collective decision that affects them’ (Parkinson 2006a; p. 151), or for Goodin:

Let us concede from the start that the ‘ideal speech situation’ would be best. The very best deliberation, let us suppose, would indeed be a cooperative game among all players in which all the deliberative virtues would be simultaneously and continuously on display.

(Goodin 2005; p. 193)

In this final section, I want to tentatively suggest a different argument on the basis of the analysis and experimental results outlined in the preceding chapters. By applying the topic of agency revision to macro accounts of deliberation, I will demonstrate that deliberative democratic systems might not be seen as ‘second-best’ on all dimensions. To do this, and following a distinction I made in chapter six, I want to differentiate my argument on the basis of two distinct levels of magnitude. The first, based on the non-robust interpretation of the data, is the (much) weaker of the two, although it is rooted in more reliable empirical foundations. The second, which emanates from the robust interpretation of the data, involves a stronger conclusion, although does so from a weaker empirical standpoint. But before I get to this, let me briefly summarise the argument I outlined in chapter two regarding the assumption, and normative

justification, of agency revision in deliberation. Using the rational choice approach, I suggested three ‘reasons’ for the favourability of democratic decision-making that triggered agency revision and caused individuals to become ‘team-reasoners’. Respectively, these justifications were as follows:

- ‘Solving’ social dilemma encounters
- Creating increased levels of trust amongst individuals
- Acting as an important element of community generation

Taken together these arguments therefore turn agency revision into an alternative dimension, other than the issue of legitimacy, upon which deliberative models of democracy can be normatively ‘judged’. As a consequence, the question that can be derived from this supposition is whether a deliberative democratic system, compared with that of unitary deliberative democracy, is likely to deliver the same positive effects. It is this final line of inquiry that I now consider.

### 7.3.1 NON ROBUST INTERPRETATION OF THE EXPERIMENTAL DATA

The first crucial point to make regarding the results from the experiments on agency revision is to look at the type of deliberation they were based upon. The use of a mini public, although structured according to four different sessions, was designed to reflect the principles associated with second-generation models of deliberation. This includes the requirement that a single group of individuals were involved in the entire process, and that all behavioural criteria were present in the single forum. In this sense, the results clearly reflect an investigation of a unitary (often termed micro) account of deliberation. Let me restate the result derived in chapter six that was based on a non-robust interpretation of the data then:

Micro Result One: Deliberation does not necessarily lead to agency revision amongst the individuals involved in the process, particularly in cases where deep disagreement amongst individuals is present.

Appropriately, this statement can be directly linked to the question of systematising deliberative democracy. In this case the experimental results relating specifically to a unitary conception of deliberation seem to suggest that the normative claim made in deliberative models of democracy does not always manifest itself. This result, crucially,

appeared to hold across all three definitions of a ‘we’ that were tested for. Team reasoners were not created in relation to the deliberating group, to the general public, and nor in relation to being a member of any deliberating group. Interpreting the data this way, it argues that no agency revision has occurred even under a unitary approach to deliberation. So long as parallel, disjointed, distributed or sequenced articulations of deliberation do not work to *reduce* the number of team-reasoners, then they must be considered no ‘worse’ than a unitary articulation on this dimension. Extending this analysis to deliberative democracy, the conclusion can be drawn that at the very least, the shift from a second to a third-generation formulation does not result automatically in a ‘second best’ alternative. As a consequence, I would put forward the following claim regarding deliberative models of democracy on this basis:

Conclusion One: on the dimension of agency revision, unitary conceptions of deliberative democracy do not *necessarily* perform better than systemic accounts of deliberative democracy.

This conclusion, it should be noted, of course raises further empirical questions. Saying something does not *necessarily* perform worse, does not mean, of course, that it might not. It simply states that without further investigation, the claim that a systematic approach to deliberative democracy is worse than a unitary approach cannot be made on the basis of agency revision.

### 7.3.2 ROBUST INTERPRETATION OF THE EXPERIMENTAL DATA

The second way of interpreting the data proceeds from the assumption that the low levels of significance were attributed to the small sample size, rather than a product of the effects of the deliberation itself. On this understanding, the results suggest that unitary forms of deliberation can indeed deliver a small degree of agency transformation, but crucially, that it appears to diminish after an initial immediate increase during the process. Indeed for games one to three (which partnered individuals with members of the same deliberative forum), the proportional increase of team-reasoners at  $t_4$ , although higher than at pre-test levels ( $t_0$ ), was significantly lower than at another time in the course of the process. In short, it appears that there might be a point in the deliberation that eventually results in a reversal of the agency revision process. Or as Shapiro (2002; p. 196) might put it, it begs the question of ‘how much is too much deliberation?’ Recall the second result that I offered in chapter six:

Micro Result Two: Deliberation can trigger a small degree of agency revision, although the effect matches an inverse ‘U’ shape relationship with respect to members of the same deliberative forum and the general public.

From this claim, it possible to draw some tentative conclusions regarding the sequenced nature of deliberation present in deliberative democratic systems. As I demonstrated in chapter five, the structure of the mini public was designed to very roughly match that of the structural account of deliberation offered in chapter one. The earlier phases, for example, were much more heavily weighted towards individuals participating in discussion focussing on education and information, whilst the latter sessions were designed to prompt more decision-focussed participation. Matching these different sessions with the experimental results, it appears to support an element of the discussion in chapter three. More specifically, in reconciling the structural account of deliberation with factors found to trigger social identity, I suggested that decision-focussed discussion, present particularly in the latter stages of a deliberation, could have the effect of priming personal identity. By making individuals openly advocate for specific preferences in cases of deep disagreement (even if these competing preferences were believed to represent a common good), it was likely that agency revision might take place in the opposite direction.

Second-generation accounts of deliberative democracy rely on a unitary conception of deliberation. All individuals affected by a decision are required to participate in all various deliberative phases with everybody else. The modest overall increase in the number of team-reasoners can therefore be used to justify support for second-generation models over wholly aggregate accounts of democracy. However, the inverse ‘U’ shape relationship within the data demonstrates that a stronger case can be made for democratic decision-making that requires individuals to only participate in an optimal amount of deliberation. Deliberative democratic systems, because they proceed on a sequenced basis, do not require all individuals to participate in a full deliberation. Indeed in Fishkin and Ackerman (2002, 2005), Goodin (2005), Hendriks (2006) and Parkinson (2006a), the only arena that requires universal participation is that which represents the public space. The formal decision-making aspect of the deliberation, including a discursive attempt at consensus, takes place in a highly restricted and small-

scale assembly. From the perspective that an approach is to be judged on the number of team-reasoners that it produces, then, it is possible to envisage a situation where deliberative democratic systems actually fare better than unitary accounts of deliberative democracy. If the vast majority of individuals are required only to participate in the early deliberative phases, then the positive agency revision effects cannot then be reversed by the more decision-focussed phases of the process. This leads to the second conclusion relating to deliberative models of democracy:

Conclusion Two: sequenced deliberation may actually result in more team-reasoners than a unitary approach. In this case and on this dimension, deliberative democratic systems might be seen as more preferable than unitary deliberative democracy.

An approach that assumes a robust interpretation of the experimental data also highlights the results of games four and five as conceptually interesting. In these two interactions, individuals were respectively partnered with a member of the general public, and another deliberating group who they had never met. In relation to the latter type of ‘we’, the number of individuals employing team reasoning increased (and stabilised) over the course of the deliberation. It did not show an inverse ‘U’ shape relationship. Recall result three that I outlined in chapter six then:

Micro Result Three: Even under conditions assumed in result two, deliberation can cause *sustained* agency revision in reference to other members of *another* deliberating group.

So what do these two results, taken together, say about different formulations of deliberation, and through that, about deliberative democratic systems more generally? Most simply, a comparison can be drawn with the notion of ‘another deliberating group’ and the parallel and disjointed articulations of deliberation discussed above. These particular accounts rely on the fact that individuals, although participating in the same overall process, do not necessarily come into contact with all the other individuals also involved. Even in disjointed deliberation, where there is some element of cross membership, the vast majority of participants do not take part in the physical action of deliberation with each other. Game five, which explicitly identified the co player as being a member of a different, but related deliberative forum, can therefore be seen as

testing this approach. Whilst the number of individuals employing team reasoning in reference to members of the same group declined towards the latter stages of the deliberation, this reduction was not seen in relation to members of another deliberating group. In other words, the fact that parallel and disjointed articulations of deliberation do not require everybody to deliberate with everybody else might actually be beneficial. To put it quite simply, they may in fact create more team-reasoners than a unitary articulation.

Linking this to the notion of a deliberative democratic system, the argument is slightly more complex than a simple transposition of the claim that it might fare better than a unitary account on this dimension. Parallel and disjointed formats of deliberation are indeed involved in the makeup of most if not all of the systematic accounts. But they have different levels of importance and formality in them. Consider the approach taken by Mansbridge (1999), Goodin (2005), Hendriks (2006) and Parkinson (2006a). The public space present in their relative deliberative democratic systems are by design, highly unstructured. Indeed, the dynamic and spontaneous nature of this arena is considered an essential feature in satisfying the democratic aspect of the decision-making process. In these types of model then, it is more likely that the individuals involved will see the other participants who they are not in contact with, as merely members of the general public. The experimental results from game four, which tested the presence of agency revision to team reasoning in respect to this group, mimics that of games one to three – an inverse ‘U’ shape relationship. On the other hand, where parallel and disjointed accounts are relied upon to deliver both deliberative and democratic elements of the system, for example in Ackerman and Fishkin’s (2002, 2005) notion of a deliberation day, the outcome is slightly different. By formalising the various forums or groups, individuals are likely to become much more aware that others are engaged in a similar deliberative process. In turn, this changes the perception of the other individuals as a member of the general public to that of a member of another deliberating group, securing more agency revision. The final tentative conclusion that I want to draw from the data then, is as follows:

Conclusion Three: parallel and disjointed deliberation may actually result in more team-reasoners than a unitary approach. In this case and on this dimension, deliberative democratic systems might be seen as more preferable than unitary deliberative democracy.

## 7.4 CONCLUDING REMARKS AND FURTHER RESEARCH

---

The various chapters of this thesis have had a number of different objectives. They have drawn on a host of different literatures and approaches to conceptually and empirically investigate the twin concepts of deliberation and deliberative democracy. Focussing on the issue of deliberative revision, I have demonstrated the conceptual distinction between the three layers of issue, preference and agency. I have then devised and undertaken empirical work in order to investigate the latter of these areas. The results, although representative of an extremely small sample size, offer some useful insights into the nature of deliberation, and in particular, the shift from a unitary to a systematic approach to deliberative democratic theory.

It is also worth noting that my results also raise some important further research questions. In relation to studying unitary deliberation at the micro level, a number of other variables can be identified as interesting for empirical investigation. For example, one particular issue to look at is the degree to which the topic under consideration impacts upon agency revision. Does deliberation cause less agency revision for individuals involved in more controversial topics, and how do the initial preferences of the individuals involved relate to this? This links specifically to the important question of whether deliberation is always appropriate for decision-making over all issues. In short, is deliberation always *better* than a wholly aggregative methodology?

Finally, and linked to the argument considered in this final chapter, I also want to highlight the possibility of more detailed investigation of the impact of specific phases in the deliberative process. My results suggest that earlier phases in the process appear to have a greater positive impact on triggering agency revision to team reasoning. Further investigation of this question, taken in conjunction with the macro articulation of a sequenced process, will have significant impact on how it is then applied in the third generation approaches discussed above. In short, I conclude this thesis by pointing out that further empirical investigation and analysis is required in order to inform normative deliberative democratic theory.



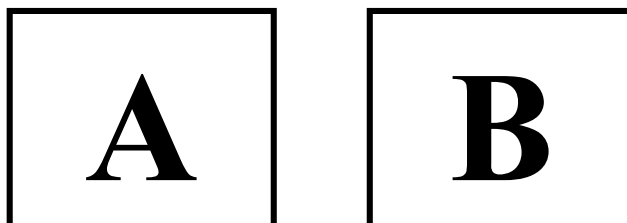
## VI. APPENDIX ONE: THE EXPERIMENT

---

### Instructions – an Example

You are playing a game against another person sat in this room. This person will be randomly assigned, and neither you nor they will know whom each other are. The outcome of the game is dependent on both yours, and your co-player's choices.

In front of you are two buttons. One is labelled "A", and the other "B".



In the game, both you and your co-player are going to press one of the two buttons.

If you both press button "A", then you get £8, and they get £8.

If you press "A", but your co-player presses "B", then you get £0, and they get £15.

If you press "B", and your co-player presses "A", then you get £15 and they get £0.

If you press "B", and your co-player presses B, then you get £5, and they get £5.

Game Choice: Please circle the button you wish to press:

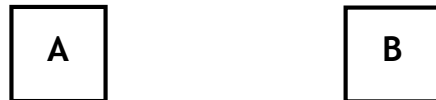


NB. Over the course of the following 5 games, you will be playing in order to collect points. At the end of the 4 discussion sessions, these points will be converted into tickets in lottery to win one of two cash prizes of £100.

## Game 1

You are playing a game against another person sat in this room. This person will be randomly assigned, and neither you nor they will know whom each other are.

In front of you are two buttons. One is labelled “A”, and the other “B”.



In the game, both you and your co-player are going to press one of the two buttons.

You Press	They Press	You Get	They Get
A	A	5 points	5 points
A	B	10 points	4 points
B	A	3 points	10 points
B	B	7 points	7 points

Game Choice: Please circle the button you wish to press:



## Game 2

You are playing a game against another person sat in this room. This person will be randomly assigned, and neither you nor they will know whom each other are.

In front of you are two buttons. One is labelled “A”, and the other “B”.



In the game, both you and your co-player are going to press one of the two buttons.

You Press	They Press	You Get	They Get
A	A	7 points	7 points
A	B	2 points	11 points
B	A	10 points	2 points
B	B	4 points	4 points

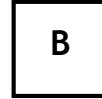
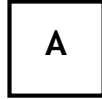
Game Choice: Please circle the button you wish to press:



### Game 3

You are playing a game against another person sat in this room. This person will be randomly assigned, and neither you nor they will know whom each other are.

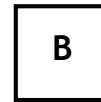
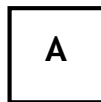
In front of you are two buttons. One is labelled “A”, and the other “B”.



In the game, both you and your co-player are going to press one of the two buttons.

You Press	They Press	You Get	They Get
A	A	10 points	10 points
A	B	20 points	2 points
B	A	2 points	16 points
B	B	15 points	15 points

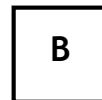
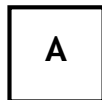
Game Choice: Please circle the button you wish to press:



### Game 4 (please note a slight difference)

You are playing a game against a **randomly selected student** from the University of York. This person will be randomly assigned.

In front of you are two buttons. One is labelled “A”, and the other “B”.



In the game, both you and your co-player are going to press one of the two buttons.

You Press	They Press	You Get	They Get
A	A	11 points	11 points
A	B	2 points	16 points
B	A	17 points	3 points
B	B	4 points	4 points

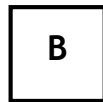
Game Choice: Please circle the button you wish to press:



### Game 5 (again please note another difference)

You are playing a game **against another person in a similar deliberating group**. This person will be randomly assigned.

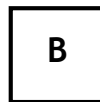
In front of you are two buttons. One is labelled “A”, and the other “B”.



In the game, both you and your co-player are going to press one of the two buttons.

You Press	They Press	You Get	They Get
A	A	4 points	4 points
A	B	20 points	2 points
B	A	3 points	20 points
B	B	13 points	13 points

Game Choice: Please circle the button you wish to press:



I hereby give permission for Thomas Flynn to use the results of these games for academic research, and more specifically, as material for submission in a PhD thesis.

Name: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

## VII. BIBLIOGRAPHY AND REFERENCES

---

- Abrams, D; Wetherall, M.; Cochrane, S.; Hogg, M., and Turner, J. (1990) "Knowing what to think by knowing who you are: Self-categorisation and the nature of norm formation, conformity and group polarization" *British Journal of Social Psychology*, vol. 29, pp. 97-119
- Ackerman, B. and Fishkin, J. (2002) "Deliberation Day" *Journal of Political Philosophy*, vol. 10, no. 2, pp. 129-152
- Ackerman, B. and Fishkin, J. (2005) *Deliberation Day* (Yale University Press)
- Ames, D. (2004) "Inside the mind reader's tool kit: Projection and stereotyping in mental state inference" *Journal of Personality and Social Psychology*, vol. 87, pp. 340-353
- Andreoni, J. (1988a) "Why free ride? Strategies and learning in public goods experiments" *Journal of Public Economics*, vol. 37, no. 3, pp. 291-304
- Andreoni, J. (1988b) "Privately provided public goods in a large economy: The limits of altruism" *Journal of Public Economics*, vol. 35, pp. 57-73
- Andreoni, J. (1990) "Impure altruism and donations to public goods: A theory of warm glow giving" *Economic Journal*, Vol. 100, pp. 464-477
- Andreoni, J. (1993) "Rational cooperation in the finitely repeated prisoners' dilemma: Experimental evidence" *Economic Journal*, vol. 103, pp. 570-585
- Andreoni, J. (1995) "Cooperation in public goods experiments: Kindness or confusion?" *American Economic Review*, vol. 84, no. 4, pp. 891-904
- Arendt, H. (1970) *On Violence* (Penguin Press)
- Aristotle [c. 347-322 BCE] (1988) *The Politics*, trans. Jowett/Barnes (Cambridge: Cambridge University Press)
- Arrow, K. (1951) *Social choice and individual values* (New York: Wiley)
- Austen-Smith, D. and Feddersen, T. (2006) "Deliberation, preference uncertainty and voting rules" *American Political Science Review*, vol. 100, no. 2, pp. 209-217
- Bacharach, M. (1993) "Variable universe games" in K. Binmore et al *Frontiers of Game Theory*, (MIT Press)
- Bacharach, M. (1999) "Interactive team reasoning: A contribution to the theory of cooperation" *Research in Economics*, vol. 53, pp. 117-147
- Bacharach, M. (2006) *Beyond individual choice: Teams and frames in game theory* (ed) N. Gold and R. Sugden (Princeton University Press)

- Bächtiger, A. and Hangartner, D. "When deliberative theory meets empirical political science: Theoretical and methodological challenges in political deliberation" *Political Studies*, vol. 58, no. 4, pp. 609-629
- Bächtiger, A.; Niemeyer, S.; Neblo, M.; Steenbergen, M. and Steiner, J. "Disentangling Diversity in deliberative democracy: Competing theories, their blind spots and complementarities" *Journal of Political Philosophy*, vol. 18, no. 1, pp. 32-63
- Banaji, M.; Hardin, C. and Rothman, A. (1993) "Implicit stereotyping in person judgement" *Journal of Personality and Social Psychology*, vol. 65, pp. 272-281
- Barber, B. (1984) *Strong Democracy* (Berkeley: University of California Press)
- Bardsley, N. and Sugden, R. (2006) "Human nature and sociality in economics" in S. Colm and J. Ythier (eds) *Handbook of the Economics of Giving, Altruism and Reciprocity Volume 1* (Elsevier)
- Barge, J. (2002) "Enlarging the meaning of group deliberation: From discussion to dialogue" in L. Frey (eds) *New directions in group communication* (Thousand Oaks CA: Sage)
- Basu, K. (2006) "Identity, trust and altruism: Sociological clues to economic development" *CAE Working Paper 06-05*
- Batson, C. (1995) "Pro social motivation: Why do we help others?" in A. Tesser (eds) *Advanced Social Psychology* (McGraw-Hill)
- Becker, G. (1976) *The economic approach to human behaviour* (University of Chicago Press)
- Beetham, D. (1991) *The Legitimation of Power* (Macmillan)
- Benhabib, S. (1996a) (eds) *Democracy and difference: Contesting the boundaries of the political* (Princeton University Press)
- Benhabib, S. (1996b) "Toward a deliberative model of democratic legitimacy" in S. Benhabib (eds) *Democracy and Difference: Contesting the Boundaries of the Political* (Princeton University Press)
- Bessette, J. (1994) *The mild voice of reason: Deliberative democracy and American national government* (Chicago: University of Chicago Press)
- Billig, M. and Tajfel, H. (1973) "Social categorisation and similarity in intergroup behaviour" *European Journal of Social Psychology*, vol. 3, pp. 27-52
- Black, L. (2008) "Deliberation, storytelling and dialogic moments" *Communication Theory*, vol. 18, pp. 93-116
- Black, D. (1948) "On the rationale of group decision-making" *Journal of Political Economy*, vol. 56, pp. 23-34

- Blair, I.; Ma, J. and Lenton, A. (2001) "Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery" *Journal of Personality and Social Psychology*, vol. 81, no. 5, pp. 828-841
- Blake, R. and Mouton, J. (1986) "From theory to practice in inter-face problem solving" in S. Worchel and W. Austin (eds) *Psychology of Intergroup Relations* (Chicago: Nelson Hall)
- Block, J. (1981) "Some enduring and consequential structures of personality" in A. Rubin, J. Aronoff, A. Barclay, and A. Zucker (eds) *Further Explorations in Psychology* (New York: Wiley)
- Bohman, J. (1995) "Public reason and cultural pluralism: Political liberalism and the problem of moral conflict" *Political Theory*, vol. 23, no. 2, pp. 253-279
- Bodenhausen, G. (1990) "Stereotypes as judgemental heuristics: Evidence of circadian variations in discrimination" *Psychological Science*, vol. 1, pp. 319-322
- Bodenhausen, G. and Lichtenstein, M. (1987) "Social stereotypes and information-processing strategies: The impact of task complexity" *Journal of Personality and Social Psychology*, vol. 52, pp. 871-880
- Bodenhausen, G.; Sheppard, L. and Kramer, G. (1994) "Negative affect and social judgement: The differential impact of anger and sadness" *European Journal of Social Psychology*, vol. 24, no. 1, pp. 45-62
- Bohman, J. (1996) *Public deliberation: Pluralism, complexity and democracy* (Cambridge: MIT Press)
- Bohman, J. (1998) "Survey article: The coming of age of deliberative democracy" *The Journal of Political Philosophy*, vol. 6, no. 4, pp. 400-425
- Bohman, J. (2009) "Epistemic value and deliberative democracy" *The Good Society*, vol. 18, no. 2, pp. 28-34
- Bohman, J. and Rehg, W. (1997) (eds) *Deliberative democracy: Essays on reason and politics* (MIT Press)
- Bohman, J. and Richardson, H. (2009) "Liberalism, deliberative democracy and reasons that all can accept" *The Journal of Political Philosophy*, vol. 17, no. 3, pp. 253-274
- Bornstein, G. (1992) "The free rider problem in intergroup conflicts over step-level and continuous public goods" *Journal of Personality and Social Psychology*, vol. 62, pp. 597-606
- Bradwell, P. (2010) *Private lives: A people's inquiry into personal information* (Demos: UK)

- Bratman, M. (1993) "Shared intentions" *Ethics*, vol. 104, pp. 97-113
- Brewer, M. (1988) "A dual process model of impression formation" in T. Srull and S. Wyer (eds) *Advances in Social Cognition* (Hillsdale: Erlbaum)
- Brewer, M. (1991) "The social self: on being the same and different at the same time" *Personality and Social Psychology Bulletin*, vol. 17, pp. 475-482
- Brewer, M. and Cramer, R. (1986) "Choice behaviour in social dilemmas: Effects of social identity, group size and decision framing" *Journal of Personality and Social Psychology*, vol. 50, no. 3, pp. 543-549
- Brewer, M. and Miller, N. (1996) *Intergroup relationships* (Buckingham, Oxford University Press)
- Brewer, M. and Harasty, A. (1996) "Seeing groups as entities: The role of perceiver motivation" in R. Sorrentino and T. Higgins (eds) *Handbook of motivation and cognition* (Vol. 3, New York: Guildford)
- Brown, R. (1988) *Group Processes* (Kent: Blackwell)
- Buchan, N.; Johnson, E. and Croson, R. (2006) "Let's get personal: An international examination of the influence of communication, culture and social distance on other regarding preferences" *Journal of Economic Behaviour and Organization*, vol. 60, pp. 373-398
- Burkhalter, S.; Gastil, J. and Kelshaw, T. (2002) "A conceptual definition and theoretical model of public deliberation in small face-to-face groups" *Communication Theory*, vol. 12, no. 4, pp. 398-422
- Campbell, D. (1958) "Common fate, similarity, and other indices of the status of aggregates of persons as social entities" *Behavioural Sciences*, vol. 3, pp. 14-25
- Catt, H. (1999) *Democracy in practice* (London: Routledge)
- Chambers, S. (1996) *Reasonable democracy: Jurgen Habermas and the politics of discourse* (Ithaca: Cornell University Press)
- Chambers, S. (2003) "Deliberative democratic theory" *Annual Review of Political Science*, vol. 6, pp. 307-326
- Chambers, S. (2004) "Behind closed doors: Publicity, secrecy and the quality of deliberation" *The Journal of Political Philosophy*, vol. 12, no. 4, pp. 389-410
- Chambers, S. (2005) "Measuring publicity's effect: Reconciling empirical research and normative theory" *Acta Politica*, vol. 40, no. 2, pp. 255-266
- Chambers, S. (2009) "Rhetoric and the public sphere: Has deliberative democracy abandoned mass democracy" *Political Theory*, vol. 37, no. 3, pp. 323-350



- Charness, G. and Ernan, H. (2002) "Altruism, fairness and reciprocity in a gift exchange experiment: An encompassing approach" *Games and Economic Behaviour*, vol. 40, pp. 203-231
- Charness, G. and Rabin, M. (2002) "Social preferences: Some simple tests and a new model" *Quarterly Journal of Economics*, vol. 117, pp. 817-869
- Chase, J. (1992) "The self and collective action: Dilemmatic identities" in G. Breakwell *Social Psychology of Identity and the Self Concept* (Surrey University Press)
- Christiano, T. (2008) *The constitution of equality* (Oxford: Oxford University Press)
- Cohen, J. (1996) "Procedure and Substance in Deliberative Democracy" in S. Benhabib (eds) *Democracy and Difference: Contesting the Boundaries of the Political* (Princeton University Press)
- Cohen, J. (1997) "Deliberation and Democratic Legitimacy" in Bohman and Rehg (1997) pp. 67-91
- Cohen, J. (1998) "Democracy and liberty" in Elster (1998a) pp. 185-231
- Cohen, J. (1999) "Reflections on Habermas on democracy" *Ratio Juris*, vol. 12, no. 4, pp. 385-416
- Cohen, J. and Rogers, J. (2003) "Power and reason" in Fung and Wright (2003).
- Coleman, S. and Gotze, J. (2001) *Bowling together: Online public engagement in policy deliberation* (London: Hansard Society)
- Colman, A.; Pulford, B. and Rose, J. (2008) "Collective rationality in interactive decisions: Evidence for team reasoning" *Acta Psychologica*, vol. 128, pp. 387-397
- Comte, A. (1983) *Auguste Comte and positivism: The essential writings* (ed) Gertrund Lenzer (University of Chicago Press)
- Cooke, M. (2000) "Five arguments for deliberative democracy" *Political Studies Review*, vol. 48, pp. 947-969
- Cookson, R. (2000) "Framing effects in public goods games" *Experimental Economics*, vol. 3, pp. 55-79
- Coote, A. and Lenaghan, J. (1997) *Citizens' Juries: Theory into practice* (IPPR)
- Costa, B. and McCrae, R. (1980) "Still stable after all these years: Personality as a key to some issues in adulthood and old age" in P. B. Baltes and O. Brim (eds) *Lifespan development and Behaviour* (New York: Academic Press)
- Cowen, T. (1993) "Altruism and the argument from offsetting transfers" in E. Paul, F. Miller and J. Paul (1993) *Altruism* (Cambridge University Press)
- Cox, J. (2004) "How to identify trust and reciprocity" *Games and Economic Behaviour*, vol. 46, pp. 260-281

- Crawford, M.; Sherman, S. and Hamilton, D. (2002) "Perceived entitativity, stereotype formation, and the interchangeability of group members" *Journal of Personality and Social Psychology*, vol. 83, no. 5, pp. 1076-1094
- Dahlberg, L. (2001) "The internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere" *Information, Communication and Society*, vol. 4, no. 4, pp. 615-633
- Darley, J. and Gross, P. (1983) "A hypothesis-confirming bias in labelling effects" *Journal of Personality and Social Psychology*, vol. 44, pp. 20-33
- Dawes, R.; McTavish, J. and Shaklee, H. (1977) "Behaviour, communication, and assumptions about other people's behaviour in a commons dilemma situation" *Journal of Personality and Social Psychology*, vol. 35, pp. 1-11
- Dawes, R., van de Kragt, M. and Orbell, J. (1988) "Not me or thee but we: The importance of group identity in eliciting cooperation in dilemma situations: Experimental manipulations" *Acta Psychologica*, vol. 68, pp. 83-97
- De Boer, J. (2008) "Collective intention, social identity, and rational choice" *Journal of Economic Methodology*, vol. 15, no. 2, pp. 169-184
- De Cramer, D. and Van Vugt, M. (1999) "Social identification effects in social dilemmas: A transformation of motives" *European Journal of Social Psychology*, vol. 29, pp. 871-893
- Delli Carpini, M.; Cook, F.; and Jacobs, L. (2004) "Public deliberation, discursive participation and citizen engagement: A review of the empirical literature" *Annual Review of Political Science*, vol. 7, pp. 315-344
- Devine, P. (1989) "Stereotypes and prejudice: their automatic and controlled components" *Journal of Personality and Social Psychology*, 56, pp. 680-690
- Dieckhoff, A. (2004) "Hebrew, the language of national daily life" in D. Judd and T. Larcone (eds) *Language, nation and state: Identity politics in a multilingual age* (Basingstoke: Palgrave Macmillan)
- Diener, D. (1977) "Deindividuation: causes and consequences" *Social Behaviour and Personality*, vol.5, no. 1, pp. 143-155
- Dijker, A. and Koomen, W. (1996) "Stereotyping and attitudinal effects under time pressure" *European Journal of Social Psychology*, vol. 26, no. 1, pp. 61-74
- Dion, K. (1973) "Cohesiveness as a determinant of ingroup-outgroup bias" *Journal of Personality and Social Psychology*, vol. 28, no. 2, pp. 163-171

- Doosje, B.; Spears, R. and Koomen, W. (1995) "When bad isn't all bad: Strategic use of sample information in generalisation and stereotyping" *Journal of Personality and Social Psychology*, vol. 69, no. 4, pp. 642-655
- Downs, A. (1957) *An economic theory of democracy* (Harper Collins)
- Drury, J; Cocking, C. and Reicher, S. (2009) "Everyone for themselves? A comparative study of crowd solidarity among emergency survivors" *British Journal of Social Psychology*, vol. 48, no. 3, pp. 487-506
- Dryzek, J. (1999) *Discursive Democracy* (Cambridge: Cambridge University Press)
- Dryzek, J. (2000) *Deliberative Democracy and Beyond: Liberals, Critics and Contestations* (Oxford University Press)
- Dryzek, J. (2001) "Legitimacy and economy in deliberative democracy" *Political Theory*, vol. 29, no. 5, pp. 651-669
- Dryzek, J. (2008) "The empirical turn in deliberative democracy", *Paper presented at the Theory and Practice of Deliberative Democracy Conference* (ANU, Canberra, 7<sup>th</sup> February)
- Dryzek, J. (2009) "Soup, society or system: Global democratisation" *Paper presented at the Deliberative Society Conference* (University of York, York, 25<sup>th</sup> June)
- Dryzek, J. (2010) "Rhetoric in democracy: A systemic appreciation" *Political Theory*, vol. 38, no. 3, pp. 319-339
- Dryzek, J. and List, C. (2003) "Social choice theory and deliberative democracy: A reconciliation" *British Journal of Political Science*, vol. 33, pp. 1-28
- Dryzek, J. and Niemeyer, S. (2006) "Reconciling pluralism and consensus as political ideals" *American Journal of Political Science*, vol. 50, no. 3, pp. 634-649
- Ellemers, N.; Wilke, H. and van Knippenberg, A. (1993) "Effects of the legitimacy of low group or individual status on individual and collective status-enhancement strategies" *Journal of Personality and Social Psychology*, vol. 64, pp. 766-778
- Elster, J. (1983) *Sour grapes: Studies in the subversion of rationality* (Cambridge University Press)
- Elster, J. (1997) "The market and the forum: Three varieties of political theory" in Bohman and Rehg (1997) pp. 3-33
- Elster, J. (1998a) *Deliberative Democracy* (Cambridge: Cambridge University Press)
- Elster, J. (1998b) "Deliberation and constitution making" in Elster (1998a) pp. 97-122
- Elster, J. (1998c) "Introduction" in Elster (1999a) pp.1-18
- Elstub, S. (2010) "The third generation of deliberative democracy" *Political Studies Review*, vol. 8, no. 3, pp. 291-307

- Epley, N.; Keysar, B. and van Boven, L. (2004) "Perspective taking as egocentric anchoring and adjustment" *Journal of Personality and Social Psychology*, vol. 87, pp. 327-339
- Escobar, O (2009) "The dialogue turn: Dialogue for deliberation", *In-Spire Journal of Law, Politics and Societies*, vol. 4, no. 2, pp. 42-70
- Estlund, D. (1997) "Beyond fairness and deliberation: The epistemic dimension of democratic authority" in J. Bohem and W. Rehg (eds) *Deliberative Democracy: Essays on Reason in Politics* (MIT Press)
- Farr, R. (1981) "The social origins of the human mind: A historical note" in J. Forgas (eds) *Social Cognition: Perspectives on Everyday Understanding* (New York: Academic Press)
- Farrar, C.; Fishkin, J.; Green, D.; List, C.; Luskin, R. and Paluck, E. (2010) "Disaggregating deliberation's effects: An experiment within a deliberative poll" *British Journal of Political Science*, vol. 40, pp. 333-347
- Fearon, J. (1998) "Deliberation as discussion" in Elster (1998a), pp. 44-68
- Fehr, E. and Schmidt, K. (2006) "The economics of fairness, reciprocity and altruism: Experimental evidence and new theories" in S. Kolm and J. Ythier (eds) *Handbook of the economics of giving, altruism and reciprocity* (Elsvier)
- Festinger, L.; Schacter, S. and Back, K. (1950) *Social pressures in informal groups: A study of a housing community* (California: Stanford University Press)
- Fisher, B. and Ellis, D. (1990) *Small group decision-making: Communication and the group process* (McGraw-Hill, New York)
- Fishkin, J. (1995) *The voice of the people* (New Haven: Yale University Press)
- Fishkin, J. (2005) "Defending Deliberation: A Comment on Ian Shapiro's The Sate of Democratic Theory" *Critical Review of International Social and Political Philosophy*, Vol. 8, No. 1, pp. 71-78
- Fishkin, J. (2009) *When the people speak: Deliberative democracy and public consultation* (Oxford: Oxford University Press)
- Fishkin, J. and Luskin, R. (1999) "Bringing deliberation to the democratic dialogue" in M. McCombs and A. Reynolds (eds) *The Poll with a Human Face: The National Issues Convention Experiment in Political Communication* (Lawrence Earlbaum)
- Fishkin, J. and Luskin, R. (2005) "Experimenting with a democratic ideal: Deliberative polling and public opinion" *Acta Politica*, vol. 40, pp. 284-298
- Fiske, T. (1998) "Stereotyping, prejudice and discrimination" in D. Gilbert, S. Fiske and G. Lindzey (eds) *The Handbook of Social Psychology* (New York: McGraw-Hill)

- Frank, R. (1988) *Passions within Reason: The Strategic Role of Emotion* (Norton)
- Freund, T.; Kruglanski, A. and Shpitzajzen, A. (1985) “The freezing and unfreezing of impression primacy: Effects of need for structure and the fear of invalidity” *Personality and Social Psychology Bulletin*, vol. 11, pp. 479-487
- Fung, A. (2007) “Mini Publics” in D. Rosenberg (eds) *Deliberation, participation and democracy: Can the people govern?* (Basingstoke: Palgrave Macmillan)
- Fung, A. and Wright, E. (2003) *Deepening democracy: Experiments in empowered participatory governance* (Princeton: Princeton University Press)
- Gastil, J. (1993) *Democracy in small groups: Participation, decision-making and communication* (Philadelphia: New Society)
- Gastil, J. (2000) *By popular demand: Revitalising representative democracy through deliberative elections* (Berkeley: University of California Press)
- Gastil, J.; Black, L.; Dees, E. and Leichter, J. (2008) “From group member to democratic citizen: How deliberating with fellow jurors reshapes civic attitudes” *Human Communication Research*, vol. 34, pp. 137-169
- Gastil, J. and Levine, P. (2005) *The Deliberative Democracy Handbook* (Jossey-Bass, San Francisco)
- Gilbert, M. (1987) “Modelling collective belief” *Synthese*, vol. 73, pp. 185-204
- Gilbert, M. (1989) *On social facts* (London: Routledge)
- Gilbert, T. and Hixon, J. (1991) “The trouble of thinking: Activation of and application of stereotypic beliefs” *Journal of Personality and Social Psychology*, vol. 60, pp. 509-517
- Gill, M. (2004) “When information does not deter stereotyping: Prescriptive stereotyping can foster bias under conditions that deter descriptive stereotyping” *Journal of Experimental Social Psychology*, vol. 40, no. 5, pp. 619-632
- Gold, N. and Sugden, R. (2007) “Collective intentions and team agency” *Journal of Philosophy*, vol. 104, no. 3, pp. 109-137
- Gold, N. and Sugden, R. (2008) “Theories of team agency” in F. Peter and S. Schmidt (eds) *Rationality and commitment* (Oxford: Oxford University Press)
- Goodin, R. (1986) “Laundering preferences” in J. Elster and A. Hylland (eds) *Foundations of social choice theory* (Cambridge: Cambridge University Press)
- Goodin, R. (2000) “Democratic deliberation within” *Philosophy and Public Affairs*, vol. 29, no. 1, pp. 81-109
- Goodin, R. (2003) *Reflective Democracy* (Oxford: Oxford University Press)

- Goodin, R. (2005) "Sequencing deliberative moments" *Acta Politica*, vol. 40, pp. 182-196
- Goodin, R. (2008a) *Innovating Democracy* (Oxford: Oxford University Press)
- Goodin, R. (2008b) "First talk, then vote" in Goodin (2008a), pp. 108-124
- Goodin, R. and Dryzek, J. (2006) "Deliberative impacts: The macro-political uptake of mini publics" *Politics and Society*, vol. 34, pp. 219-244
- Goodin, R. and Niemeyer, S. (2003) "When does deliberation begin? Internal reflection versus public discussion in deliberative democracy" *Political Studies*, vol. 51, no. 4
- Green, D. and Shapiro, I. (1994) *Pathologies of rational choice theory: A critique of applications in political science* (Yale University Press)
- Gunderson, A. (1995) *The environmental promise of democratic deliberation* Madison: University of Wisconsin Press)
- Guston, G. (1999) "Evaluating the first US Consensus Conference: The impact of the citizens' panel on telecommunications and the future of democracy" *Science, Technology and Human Values*, Vol. 24, pp. 451-482
- Gutmann, A. (1999) "Deliberative democracy and majority rule" in H. Hongju Koh and R. Slye (eds) *Deliberative Democracy and Human Rights* (New Haven: Yale University Press)
- Gutmann, A. and Thompson, D. (1996) *Democracy and Disagreement* (Cambridge: Harvard University Press)
- Gutmann, A. and Thompson, D. (2004) *Why Deliberative Democracy* (Princeton: Princeton University Press)
- Habermas, J. (1987) *The Theory of Communicative Action*, vol. 2, trans. T. McCarthy (Boston: Beacon Press)
- Habermas, J. (1990) *Moral Consciousness and Communicative Action* (MIT Press)
- Habermas, J. (1994) "Three normative models of democracy" *Constellations*, vol. 1, no. 1, pp. 1-10
- Habermas, J. (1995) "Reconciliation through the Public Use of Reason: Remarks on John Rawls's Political Liberalism" *Journal of Philosophy*, Vol. 17, No. 3
- Habermas, J. (1996a) "Three Normative Models of Democracy" in S. Benhabib *Democracy and Difference: Contesting the Boundaries of the Political* (Princeton University Press)
- Habermas, J. (1996b) *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (MIT Press)

- Habermas, J. (1997) "Popular Sovereignty as Procedure" in Bohman and Rehg (1997); pp. 35-65
- Habermas, J. (1999) "Introduction" *Ratio Juris*, vol. 12, no. 4, pp. 329-335
- Haldane, J. (1932) *The causes of evolution* (Princeton University Press)
- Hamilton, D. and Sherman, J. (1994) "Stereotypes" in R. Wyer and T. Srull (eds) *Handbook of Social Cognition* (Hillsdale: Erlbaum)
- Hamilton, D. and Sherman, S. (1996) "Perceiving persons and groups" *Psychological Review*, vol. 103, no. 2, pp. 336-355
- Hamilton, D. and Trolie, T. (1986) "Stereotypes and stereotyping: An overview of the cognitive approach" in J. Dovidio and S. Gaertner (eds) *Prejudice, discrimination and racism* (Orlando: Academic Press)
- Hamlett, P. and Cobb, D. (2006) "Potential solutions to public deliberation problems: Structured deliberations and polarisation cascades" *The Policy Studies Journal*, vol. 34, no. 4, pp. 629-648
- Harsanyi, J. (1955) "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility" *Journal of Political Economy*, vol. 63, pp. 309-321
- Haslam, S.; Oakes, P.; Reynolds, K. and Turner, J. (1999) "Social identity salience and the emergence of stereotype consensus" *Personality and Social Psychology Bulletin*, vol. 25, no. 7, pp. 809-818
- Haugtvedt, C. and Petty, R. (1992) "Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes" *Journal of Personality and Social Psychology*, vol. 63, pp. 308-319
- Heilman, M. (2001) "Description and prescription: How gender stereotypes prevent women's ascent up the organisational ladder" *Journal of Social Issues*, vol. 57, no. 4, pp. 657-674
- Hendricks, C. (2005) "Consensus conferences and planning cells" in J. Gastil and P. Levine (eds) *The Deliberative Democracy Handbook* (Jossey-Bass, San Francisco)
- Hendricks, C. (2006) "Integrated deliberation: Reconciling civil society's dual roles in deliberative democracy" *Political Studies*, vol. 54, pp. 486-508
- Hendricks, C. (2009) "The democratic soup: Mixed meanings of representation in governance networks" *Governance: An International Journal of Policy, Administration and Institutions*, vol. 22, no. 4, pp. 689-715
- Heritage, J. and Clayman, S. (2010) *Talk in Action: Interactions, Identity and Institutions* (Wiley Blackwell)

- Hodgson, D. (1967) *Consequences of utilitarianism* (Clarendon Press)
- Hoffman, E.; McCabe, K.; Shachat, K. and Smith, V. (1992) "Preferences, property rights, and anonymity in bargaining games" *Games and Economic Behaviour*, vol. 7, pp. 346-380
- Hogg, M. (1992) *The Social Psychology of Group Cohesiveness: From Attraction to Social Identity* (Harvester Wheatsheaf)
- Hogg, M. and McGarty, C. (1990) "Self-categorisation and social identity" in D. Abrams and M. Hogg (eds) *Social Identity Theory: Constructive and critical advances* (New York: Springer-Verlag)
- Hogg, M. and Turner, J. (1987) "Intergroup behaviour, self-stereotyping and the salience of social categories" *British Journal of Social Psychology*, vol. 26, pp. 325-340
- Hogg, M. and Vaughan, G. (2005) *Social psychology: 4<sup>th</sup> Edition* (Prentice Hall)
- Hollis, M. (1998) *Trust within reason* (Cambridge University Press)
- Hurley, S. (1989) *Natural reasons* (Oxford: Oxford University Press)
- James, K. (1993) "Conceptualising self with in-group stereotypes: Context and esteem precursors" *Personality and Social Psychology Bulletin*, vol. 19, pp. 117-121
- Janssen, D. and Kies, R. (2005) "Online forums and deliberative democracy" *Acta Politica*, vol. 40, pp. 317-335
- Jussim, L.; Nelson, T.; Manis, T. and Sofin, S. (1995) "Prejudice, stereotypes, and labelling effects: Sources of bias in person perception" *Journal of Personality and Social Psychology*, vol. 68, pp. 228-246
- Kagel, J. and Wolfe, K. (2001) "Tests of fairness models bases on equity considerations in a three-person ultimatum game" *Experimental Economics*, vol. 4, pp. 203-219
- Kahneman, D. and Tversky, A. (1981) "The framing of decisions and the psychology of choice", *Science*, vol. 211, pp. 453-458
- Kelley, H. (1952) "Two functions of reference groups" in G. Swanson, T. Newcomb and E. Hartley (eds) *Reading in Social Psychology* (New York: Holt, Rinehart and Winston)
- Knight, J. and Johnson, J. (1994) "Aggregation and Deliberation: On the Possibility of Democratic Legitimacy" *Political Theory*, Vol. 22, pp. 279-296
- Koole, S.; Smeets, K.; Van Knippenberg, A. and Dijksterhuis, A. (1999) "The cessation of rumination through self-affirmation" *Journal of Personality and Social Psychology*, vol. 77, no.1, pp. 111-125



- Kramer, R. and Brewer, M. (1984) "Effects of group identity on resource use in a simulated commons dilemma" *Journal of Personality and Social Psychology*, vol. 46, no. 5, pp. 1044-1057
- Krause, S. (2008) *Civil passions: Moral sentiment and democratic deliberation* (Princeton: Princeton University Press)
- Krueger, J. (2007) "From social projection to social behaviour" *European Review of Social Psychology*, vol. 18, no. 1, pp. 1-35
- Krueger, J. and Acevedo, M. (2007) "Perceptions of the self and other in the prisoner's dilemma: Outcome bias and evidential reasoning" *American Journal of Psychology*, vol. 120, no. 4, pp. 593-618
- Krueger, J. and Stanke, D. (2001) "The role of self-referent and other-referent knowledge in perceptions of group characteristics" *Personality and Social Psychology Bulletin*, vol. 27, pp.878-888
- Kruglanski, A. and Freund, T. (1983) "The freezing and unfreezing of lay inferences: Effects on impression primacy, ethnic stereotyping and numerical anchoring" *Journal of Experimental Social Psychology*, vol. 19, pp. 448-468
- Kunda, Z. and Oleson, K. (1995) "Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants" *Journal of Personality and Social Psychology*, vol. 68, pp. 565-579
- Kunda, Z. and Thagard, P. (1996) "Forming impressions from stereotypes, traits and behaviours: A parallel constraint satisfaction theory" *Psychological Review*, vol. 103, pp. 284-308
- Lang, A. (2007) "But is it for real? The British Columbia Citizens Assembly as a model of state sponsored citizen empowerment" *Politics and Society*, vol. 35, pp. 35-69
- Lave, L. B. (1965) "Factors affecting cooperation in the prisoner's dilemma" *Behavioural Science*, vol. 10, pp. 26-38
- Ledyard, R. (1995) "Public goods experiments: A review" in A. Roth and J. Kagel (eds) *The Handbook of Experimental Economics* (Princeton University Press)
- Leib, E. (2006) "Can direct democracy be made deliberative?" *University of Buffalo Law Review*
- Levy, B. (1996) "Improving memory in old age through implicit self-stereotyping" *Journal of Personality and Social Psychology*, vol. 71, pp. 1092-1107
- List, C. and Dietrich, F. (forthcoming) "A model of non-informational preference change" *Journal of Theoretical Politics*
- List, C.; Luskin, R.; Fishkin, J. and McLean, I. (2006) "Deliberation, single-peakedness,

- and the possibility of meaningful democracy: evidence from deliberative polls”  
*PSPE Working Paper Series*, 01-2006, LSE, UK
- Locksley, A.; Oritz, V. and Hepburn, C. (1980) “Social categorisation and discriminatory behaviour: Extinguishing the minimal intergroup discrimination effect” *Journal of Personality and Social Psychology*, vol. 39, pp. 773-783
- Loomes, G.; Starmer, C. and Sugden, R. (2003) “Do anomalies disappear in repeated markets” *Economic Journal*, vol. 113, pp. 153-166
- Loomes, G. and Sugden, R. (1982) “Regret theory: An alternative model of rational choice under uncertainty” *Economic Journal*, vol. 92, pp. 805-824
- Lowenstein, G. (1999) “Experimental economics from the vantage-point of behavioural economics” *The Economic Journal*, vol. 109, pp. F25-F34
- Luskin, R., Fishkin, J. and Park, A. (1999) “Learning and voting in Britain: insights from the Deliberative Poll”, paper presented at the annual meeting of the APSA, Atlanta
- Luskin, R., Fishkin, J., McAllister, I., Higley, J. and Ryan, P. (2000) “Information effects in referendum voting: Evidence from the Australian Deliberative Poll”, paper presented at the annual meeting of the APSA, Washington DC
- Luskin, R., Iyengar, S. and Fishkin, J. (2004) “Considered opinions on U.S foreign policy: Face to face versus online Deliberative Polling” Research Paper, *Centre for Deliberative Democracy* (Stanford: Stanford University)
- Macrae, N.; Milne, A. and Bodenhausen, G. (1994) “Stereotypes as energy saving devices: A peek inside the cognitive toolbox” *Journal of Personality and Social Psychology*, vol. 66, pp. 37-47
- Makau, J. and Marty, D. (2001) *Cooperative argumentation: A model for a deliberative community* (Illinois, Waveland)
- Manin, B.; Stein, E. and Mansbridge, J. (1987) "On Legitimacy and Political Deliberation" *Political Theory*, Vol. 15, No. 3, pp. 338-368
- Mansbridge, J. (1980) *Beyond Adversarial Democracy* (University of Chicago Press)
- Mansbridge, J. (1990) “Feminism and democracy” *American Prospect*, vol. 1, pp. 126-139
- Mansbridge, J. (1999) “Everyday talk in the deliberative system” in S. Macedo (eds) *Deliberative politics: Essays on democracy and disagreement* (New York: Oxford University Press)
- Mansbridge, J. (2006) “Conflict and self interest in deliberation” in S. Besson and J. Marti (eds) *Deliberative democracy and its discontents* (Aldershot: Ashgate)

- Mansbridge, J. (2010a) "The deliberative system disaggregated" *Paper presented at the 2010 APSA Annual Meeting*
- Mansbridge, J. (2010b) "Deliberative polling as the gold standard" *The Good Society*, vol. 19, no. 1, pp. 55-62
- Mansbridge, J.; Bohman, J.; Chambers, S.; Estlund, D.; Follesdai, A.; Fung, A.; Lafont, C.; Manin, B. and Marti, J.L. (2010) "The place of self-interest and role of power in deliberative democracy" *The Journal of Political Philosophy*, vol. 18, no. 1, pp. 64-100
- Margolis, H. (1981) "A new model of rational choice" *Ethics*, vol. 91, pp. 265-279
- Margolis, H. (1984) *Selfishness, altruism and rationality: A theory of social choice* (Chicago: Chicago University Press)
- Markova, I. (1987) *Human awareness* (London: Hutchinson)
- Markovits, E. (2006) "The trouble with being earnest: Deliberative democracy and the sincerity norm" *The Journal of Political Philosophy*, vol. 14, no. 3, pp. 249-269
- Markus, H. (1977) "Self-schemata and processing information about the self" *Journal of Personality and Social Psychology*, vol. 35, no. 1, pp. 63-78
- Markus, H. and Nurius, P. (1986) "Possible selves" *American Psychologist*, vol. 41, no. 9, pp. 954-96
- Markus, H. and Wurf, E. (1987) "The dynamic self-concept: A social psychological perspective" *Annual Review of Psychology*, vol. 38, pp. 299-337
- Marwell, G. and Ames, R. (1981) "Economists free ride, does anyone else? Experiments in the provision of public goods, part IV" *Journal of Public Economics*, vol. 15, pp. 295-310
- McClintock, C. (1972) *Experimental social psychology* (eds) (Holt, Rinehart and Winstone)
- McKillip, J.; Dimicelli, A. and Luebke, L. (1977) "Group salience and stereotyping" *Social Behaviour and Personality*, vol. 5, no. 1, pp. 81-85
- McLeod, M.; Scheufele, D.; Moy, P.; Horowitz, P.; Holbert, E.; Zhang, R.; Zubric, W.; and Zubric, J (1999) "Understanding deliberation: The effects of discussion networks on participation in public forum" *Communication Research*, vol. 26, pp. 743-774
- Mehta, J.; Starmer, C. and Sugden, R. (1994) "The nature of salience: an experimental investigation of pure coordination games" *The American Economic Review*, vol. 84, no.3, pp. 658-673

- Mendelberg, T. (2002) "The deliberative citizen: theory and evidence" in M. Delli Carpini, L. Huddy and R. Shapiro (eds) *Research in micropolitics: Political decision-making, deliberation and participation* (Greenwich CT, JAI Press)
- Mill, J.S. (1972) *Utilitarianism, On Liberty, Considerations on Representative Government*, edited by H. B. Acton (Everyman)
- Miller, D. (1992) "Deliberative democracy and social choice" *Political Studies*, vol. 40, pp. 54-67
- Miller, P and Rose, N. (2008) *Governing the Present* (Polity Press, Cambridge)
- Minozzi, W.; Neblo, M. and Siegel, D. (2010) "A theory of deliberation as interactive reasoning" Paper presented to the 2010 Annual Meeting of the Midwest Political Science Association, Chicago.
- Mullen, B.; Brown, N. and Smith, C. (1992) "Ingroup bias as a function of salience, relevance and status: An integration" *European Journal of Social Psychology*, vol. 22, pp. 103-122
- Neblo, M. (2007a) "Family disputes: Diversity in defining and measuring deliberation" *Swiss Political Science Review*, vol. 13, no. 4, pp. 527-557
- Neblo, M. (2007b) "Change for the better? Linking the mechanisms of deliberative opinion change to normative theory" *Working Paper available at <http://polisci.osu.edu/faculty/mneblo/papers.htm>*
- Nelson, T.; Acker, M. and Manis, M. (1996) "Irrepressible stereotypes" *Journal of Experimental Social Psychology*, vol. 32, no. 1, pp. 13-38
- Niemeyer, S. (2004) "Deliberation in the wilderness: Displacing symbolic politics" *Environmental Politics*, vol. 13, no. 2, pp. 347-372
- Niemeyer, S. and Dryzek, J. (2007) "The ends of deliberation: Meta-consensus and inter-subjective rationality as ideal outcomes" *Swiss Political Science Review*, vol. 13, no. 4, pp. 497-526
- Nolan, M.; Haslam, S; Spears, R and Oakes, P. (1999) "An examination of resource-based and fit-based theories of stereotyping under cognitive load and fit" *European Journal of Social Psychology*, vol. 29, pp. 641-663
- Oakes, P. (1987) "The salience of social categories" in J. Turner, M. Hogg, P. Oakes, S. Reicher and M. Wetherell (eds) *Rediscovering the social group: A self categorisation theory* (Oxford: Blackwell)
- Oakes, P.; Haslam, S. and Turner, J. (1994) *Stereotyping and social reality* (Oxford: Blackwell)

- Oakes, P. and Turner, J. (1986) "Distinctiveness and the salience of social category memberships: is there an automatic perceptual bias towards novelty?" *European Journal of Social Psychology*, vol. 16, pp. 325-344
- Olson, M. (1965) *The logic of collective action* (Cambridge: Harvard University Press)
- Orbell, J.; van de Kragt, A.; and Dawes, R. (1988) "Explaining discussion induced cooperation" *Journal of Personality and Social Psychology*, vol. 54, pp. 811-819
- Ozanne, J., Corus, C. and Saatcioglu, B. (2009) "The philosophy and methods of deliberative democracy: Implications for public policy and marketing" *Journal of Public Policy and Marketing*, Vol. 28, pp. 29-40
- Page, B. (1996) *Who deliberates?* (Chicago: University of Chicago Press)
- Parkinson, J. (2003) "Legitimacy problems in deliberative democracy" *Political Studies*, vol. 51, no. 1, pp. 180-196
- Parkinson, J. (2006a) *Deliberating in the real world: Problems of legitimacy in deliberative democracy* (Oxford: Oxford University Press)
- Parkinson, J. (2006b) "Rickety Bridges: Using the Media in Deliberative Democracy" *British Journal of Political Science*, Vol. 36, pp. 175-183
- Parkinson, J. (2007) "The house of lords: A deliberative democratic defence" *Political Quarterly*, vol. 73, no. 3, pp. 374-381
- Pearce, W. and Littlejohn, S. (1997) *Moral conflict: When social worlds collide* (California: Sage)
- Perdue, C.; Dovidio, J.; Gurtman, M.; and Richard, B. (1990) "Us and Them: Social categorisation and the process of intergroup bias" *Journal of Personality and Social Psychology*, vol. 59, no. 3, pp. 475-486
- Pettit, P. and Schweikard, D. (2006) "Joint actions and group agents" *Philosophy of Social Sciences*, vol. 36, no. 1, pp. 18-39
- Petty, R.; Haugtvedt, C. and Smith, S. (1995) "Elaboration as a determinant of attitude strength: Creating attitudes that are persistent, resistant and predictive of behaviour" in R. Petty and J. Krosnick (eds) *Attitude strength: Antecedents and consequences* (Mahwah: Erlbaum)
- Petty, R. and Krosnick, J. (1995) (eds) *Attitude strength: Antecedent and consequences* (Mahwah Earlbaum)
- Platt, M. (2008) "Participation for what? A policy-motivated approach to political activism" *Political Behaviour*, vol. 30, no. 3, pp. 391-413
- Prentice, D. and Miller, D. "The psychology of ingroup attachment" Paper presented at the conference on The Self and the Collective, Princeton University

- Przeworski, A. (1998) "Deliberation and Ideological Domination" in Elster (1998a) pp. 140-160
- Rabbie, J. and Horwitz, M. "Arousal of ingroup-outgroup attachment by a chance win or loss" *Journal of Personality and Social Psychology*, vol. 13, no. 3, pp. 269-277
- Rabin, M. (1993) "Incorporating fairness into game theory and economics" *American Economic Review*, vol. 83, pp. 1281-1302
- Rawls, J. (1993) *Political Liberalism* (New York: Columbia University Press)
- Rawls, J. (1997a) "The Idea of Public Reason" in J. Bohman and W. Rehg *Deliberative Democracy: Essays on Reason in Politics* (MIT Press)
- Rawls, J. (1997b) "The Idea of Public Reason Revisited" *University of Chicago Law Review*, vol. 94, pp. 765-807
- Regan, D. (1980) *Utilitarianism and cooperation* (Clarendon Press)
- Riker, W. (1988) *Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice* (Waveland Pr)
- Riker, W. (1982) *Liberalism against populism* (San Francisco: W H Freeman)
- Risse, T. (2000) "Let's argue: Communicative action in world politics" *International Organisation*, vol. 54, pp. 1-39
- Roth, A. (1995) "Bargaining experiments" in J. Kagel and A. Roth (eds) *The Handbook of Experimental Economics* (Princeton: Princeton University Press)
- Rudman, L. and Glick, P. (2001) "Prescriptive gender stereotypes and backlash toward agentic women" *Journal of Social Issues*, vol. 57, pp. 743-762
- Ryan, C. and Bogart, L. (2001) "Longitudinal changes in the accuracy of new group members' in-group and out-group stereotypes" *Journal of Experimental Social Psychology*, vol. 37, pp. 118-133
- Rydell, R.; Hugenberg, H.; Ray, D. and Mackie D. (2007) "Implicit theories about groups and stereotyping: The role of group entitativity" *Personality and Social Psychology Bulletin*, vol. 33, pp. 549-558
- Ryfe, D. (2005) "Does deliberative democracy work?" *Annual Review of Political Science*, vol. 8, pp. 49-71
- Ryfe, D. (2006) "Narrative and deliberation in small group forums" *Journal of Applied Communication Research*, vol. 43, no. 1, pp. 72-93
- Sanders, L. (1997) "Against deliberation" *Political Theory*, vol. 25, no. 3, pp. 347-376
- Savage, L. (1954) *The foundations of statistics* (New York: Wiley)
- Saward, M. (1998) *The terms of democracy* (Cambridge: Polity Press)

- Saward, M. (2006) "Representation" in A. Dobson, and R. Eckerson (eds) *Political Theory and the Ecological Challenge* (Cambridge: Cambridge University Press)
- Schelling, T. (1960) *The strategy of conflict* (Cambridge: Harvard University Press)
- Schkade, D.; Sunstein, C. and Hastie, R. (2007) "What happened on deliberation day?" *California Law Review*, vol. 95, pp. 915-940
- Schlenker, B. (1985) *The self and social life* (New York: McGraw Hill)
- Schwartzman, M. (2004) "The completeness of public reason" *Politics, Philosophy and Economics*, vol. 3, no. 2, pp. 191-220
- Searle, J. (1990) "Collective intentions and actions" in P. Cohen, J. Morgan and M. Pollack (eds) *Intentions in Communication* (MIT Press)
- Sen, A. (1966) "A Possibility Theorem of Majority Decisions" *Econometrica*, vol. 34, pp. 491-499
- Sen, A. (1992) *Inequality Re-examined* (Cambridge: Harvard University Press)
- Sesko, A. and Biernat, M. (2010) "Prototypes of race and gender: The invisibility of black women" *Journal of Experimental Social Psychology*, vol. 46, no. 2, pp. 356-360
- Shapiro, I. (2002) "Optimal deliberation" *The Journal of Political Philosophy*, vol. 10, no. 2, pp. 196-211
- Shapiro, I. (2003) *The State of Democratic Theory* (Princeton University Press)
- Sherif, M.; Harvery, O.; White, B.; Hood, W.; and Carolyn, W. (1961) *Intergroup conflict and cooperation: The robbers cave experiment* (EBook available at [psychclassics.yorku.ca/sherif](http://psychclassics.yorku.ca/sherif))
- Sherman, J. (1996) "Development and mental representation of stereotypes" *Journal of Personality and Social Psychology*, vol. 70, no. 6, pp. 1126-1141
- Sherman, S.; Hamilton, D. and Lewis, A. "Perceived entativity and the social identity value of group memberships" in D. Abrams and M. Hogg (eds) *Social Identity and Social Cognition* (Oxford: Blackwell)
- Simon, B. and Hamilton, D. (1994) "Self Stereotyping and Social Context: The effects of Relative in-group size and in-group status" *Journal of Personality and Social Psychology*, vol. 66, no.4 pp. 699-711
- Sinclair, S.; Hardin, C. and Lowery, B. (2006) "Self-stereotyping in the context of multiple social identities" *Journal of Personality and Social Psychology*, vol. 90, no. 4, pp. 529-542
- Smith, V. (1962) "An experimental study of competitive market behaviour" *Journal of Political Economy*, vol. 70, pp. 111-137

- Smith, A. (1976b)[1776] *An inquiry into the nature and causes of the wealth of nations* (OUP)
- Smith, A. (1976a) [1759] *The theory of moral sentiments* (OUP)
- Smith, G. (2009) *Democratic Innovations* (CUP)
- Smith, E. and Henry, S. (1996) “An in-group becomes part of the self: Response time evidence” *Personality and Social Psychology Bulletin*, vol. 22, pp. 635-64
- Smith, A.; Miller, D.; Maltner, A.; Crump, S.; Garcia-Marques, T. and Mackie, D. (2005) “Familiarity can increase stereotyping” *Journal of Experimental Social Psychology*, vol. 42, no. 4, pp. 471-478
- Smith, G.; Wales, C.; John, P.; Cotterill, S.; Sturgis, P.; Stoker, G. and Namura, H. (2009) “Measuring the ‘deliberative quality’ of an online experimental mini-public: Methodology and early results” Paper presented to the ECPR General Conference, Postdam September 2009
- Spencer-Rodgers, J.; Hamilton, D. and Sherman, S. (2007) “The central role of entitativity in stereotypes of social categories and task groups” *Journal of Personality and Social Psychology*, vol. 92, no. 3, pp. 369-388
- Sporndli, M. (2003) “Discourse quality and political decisions: An empirical analysis of debates in the German conference committee” *Discussion Paper SP IV 2003-101*, Social Science Research Center Berlin
- Starmer, C. (2000) “Developments in non expected utility theory: The hunt for a descriptive theory of choice under risk” *Journal of Economic Literature*, vol. 38, pp.332-382
- Steenbergen, M; Bachtiger, B.; Sporndli, M. and Steiner, J. (2003) “Measuring political deliberation: A discourse quality index” *Comparative European Politics*, vol. 1, pp. 21-48
- Steiner, J. (2008) “Concept stretching: The case of deliberation” *European Political Science*, vol. 7, pp. 186-190
- Steiner, J.; Bachtiger, A.; Sporndli, M. and Steenbergen, M. (2004) *Deliberative politics in action* (Cambridge: Cambridge University Press)
- Stewart, T.; Weeks, M. and Lupfer, M. (2003) “Spontaneous stereotyping: A matter of prejudice?” *Social Cognition*, vol. 21, no. 4, pp. 263-298
- Sugden, R. (1984) “Consistent conjectures and voluntary contributions to public goods: Why conventional theory does not work” *Journal of Public Economics*, vol. 27, pp. 117-2



- Sugden, R. (1993) "Thinking as a team: Towards an explanation of non-selfish behaviour in E. Paul, F. Miller and J. Paul (1993) *Altruism* (Cambridge University Press)
- Sugden, R. (1995) "A theory of focal points" *Economic Journal*, vol. 105, pp. 533-550
- Sugden, R. (2000) "Team preferences" *Economics and Philosophy*, vol. 16, pp. 175-204
- Sugden, R. (2002) "Beyond sympathy and empathy: Adam Smith's concept of fellow feeling" *Economics and Philosophy*, vol. 18, pp. 63-87
- Sugden, R. (2003) "The logic of team reasoning" *Philosophical Explorations*, vol. 6, pp. 165-181
- Sugden, R. (2006) "Taking unconsidered preferences seriously" *Philosophy*, Supplement 59, pp. 209-232
- Sugden, R. (2008) "Nash equilibrium, team reasoning and cognitive hierarchy theory" *Acta Psychologica*, vol. pp. 402-408
- Suls, J. (1982) (ed) *Psychological perspectives on the self* (Hillsdale: Erlbaum)
- Sunstein, C. (1996) "Leaving things undecided" *Harvard Law Review*, vol. 110, pp. 4-101
- Sunstein, C. (1997) "Deliberation, democracy, disagreement" in R. Bontekoe and M. Stepaniants (eds) *Justice and democracy: Cross-cultural perspectives* (Honolulu: University of Hawaii Press)
- Sunstein, C. (2002) "The law of group polarisation" *The Journal of Political Philosophy*, vol. 10, no. 2, pp. 175-195
- Swann, W. and Hill, C. (1982) "When our identities are mistaken: Reaffirming self-Concepts through Social Interaction" *Journal of Personality and Social Psychology*, vol. 43, pp. 59-66
- Tajfel, H. (1970) "Experiments in intergroup discrimination" *Scientific American*, vol. 223, pp. 96-102
- Tajfel, H. (1974) "Social identity and intergroup behaviour" *Social Science Information*, vol. 13, pp. 65-93
- Tajfel, H. (1978) (ed) *Differentiation between social groups* (Cambridge, UK: CUP)
- Tajfel, H. (1982) (ed) *Social identity and intergroup relations* (Cambridge University Press)
- Tajfel, H.; Billig, M.; Bundy, R. and Flament, C. (1979) "Social categorisation and intergroup behaviour" *European Journal of Social Psychology*, vol. 1, pp. 149-177

- Tajfel, H. and Turner, J. (1986) "The social identity theory of intergroup behaviour" in S. Worchel and W. Austin (eds) *Psychology of Intergroup Relations* (Nelson-Hall)
- Tannen, D. (1998) *The argument culture: Moving from debate to dialogue* (New York: Random House)
- Taylor, D. and Dube, L. (1986) "Two faces of identity: The 'I' and the 'we'" *Journal of Social Issues*, vol. 42, no. 2, pp. 81-98
- Thaler, R. (1987) "The psychology of choice and the assumptions of economics" in A. Roth (eds) *Laboratory Experimentation in Economics: Six Points of View* (Cambridge University Press)
- Thompson, D. (2008) "Deliberative democratic theory and empirical political science" *Annual Review of Political Science*, vol. 11, pp. 497-520
- Thurstone, L. (1931) "The indifference function" *Journal of Social Psychology*, vol. 2, pp. 139-167
- Turner, J. (1981) "The experimental social psychology of intergroup behaviour" in J. Turner and H. Giles (eds) *Intergroup Behaviour* (Oxford: Blackwell)
- Turner, J. (1982) "Towards a cognitive redefinition of the group" in H. Tajfel (eds) *Social Identity and Intergroup Relations*, (Cambridge: Cambridge University Press)
- Turner, J. (1984) "Social identification and psychological group formation" in H. Tajfel (eds) *The Social Dimension: European Developments in Social Psychology* (vol. 2) (Cambridge: Cambridge University Press)
- Turner, J. (1985) "Social identification and the self concept: A cognitive theory of group behaviour" in E. Lawler (eds) *Advances in group processes: Theory and research* (Greenwich: JAI Press)
- Turner, J.; Hogg, M.; Oakes, P.; Reicher, S. and Wetherell, M. (1987) *Rediscovering the social group: A self-categorisation theory* (Blackwell)
- van Lange, P. (1999) "The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation" *Journal of Personality and Social Psychology*, vol. 77, no. 2, pp. 337-349
- van Stokkom, B. (2003) "Deliberative group dynamics: Power, status and affect in interactive policy making" *Policy and Politics*, vol. 33, pp. 387-409
- von-Neumann, J. and Morgenstern, O. (1944) *The theory of games and economic behaviour* (Princeton University Press)

- Wallis, W. and Friedman, M. (1942) "The empirical derivation of indifference functions" in O. Lange, F. McIntyre and T.O. Yntema (eds) *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz* (University of Chicago Press)
- Warren, M. (1992) "Democratic theory and self-transformation" *American Political Science Review*, vol. 86, pp. 8-23
- Warren, M. and Pearse, H. (2008) *Designing deliberative democracy* (CUP)
- Weber, M. (1968) [1922] *Economy and society* (University of California Press: Berkley)
- Wegener, D.; Clark, J. and Petty, R. (2006) "Not all stereotyping is created equal: Differential consequences of thoughtful versus non-thoughtful stereotyping" *Journal of Personality and Social Psychology*, vol. 90, no. 1, pp. 42-59
- Wilder, S. and Thompson, J. (1980) "Intergroup contact with independent manipulation of in-group and out-group interaction" *Journal of Personality and Social Psychology*, vol. 38, pp. 589-603
- Wilson, W. and Katayani, M. (1968) "Intergroup Attitudes and strategies in games between opponents of the same or different race" *Journal of Personality and Social Psychology*, vol. 9, pp. 24-30
- Wollheim, R. (1962) "A Paradox in the Theory of Democracy" in P. Laslett and W. Runciman *Philosophy, Politics and Society* (eds, 2<sup>nd</sup> edition) (Oxford, Blackwell)
- Young, I. (1996) "Communication and the other: Beyond deliberative democracy" in S. Benhabib (eds) *Democracy and Difference: Contesting the boundaries of the political* (Princeton: Princeton University Press)
- Young, I. (1999) "Justice, inclusion and deliberative democracy" in S. Macedo (eds) *Deliberative Politics: Essays on democracy and disagreement* (Oxford: Oxford University Press)
- Young, I. (2001) "Activist challenges to deliberative democracy" *Political Theory*, vol. 29, no. 5, pp. 670-690