

**Byron Rienstra and [Derek Hook](#)**  
**Weakening Habermas : the undoing of  
communicative rationality**

**Article (Accepted version)  
(Refereed)**

**Original citation:**

Rienstra, Byron and Hook, Derek (2006) Weakening Habermas : the undoing of communicative rationality. *Politikon: South African journal of political studies*, 33 (3). pp. 313-339.

DOI: [10.1080/02589340601122950](https://doi.org/10.1080/02589340601122950)

© 2006 [Taylor & Francis](#)

This version available at: <http://eprints.lse.ac.uk/955/>

Available in LSE Research Online: June 2008

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final manuscript accepted version of the journal article, incorporating any revisions agreed during the peer review process. Some differences between this version and the published version may remain. You are advised to consult the publisher's version if you wish to cite from it.

## WEAKENING HABERMAS: THE UNDOING OF COMMUNICATIVE RATIONALITY

Byron Rienstra \* and Derek Hook \*\*

*ABSTRACT Abstract: Habermas's elaboration of a procedural, discursive deliberative democracy extends from his faith in communicative action, in symmetrical communicative interactions played out in an arena of communicative rationality. Yet Habermas expects too much of his agents. His theory of communicative action, built upon the necessary possession of communicative rationality, requires individuals to have clear, unfettered access to their own reasoning, possessing clear preference rankings and defensible rationales for their goals and values. Without such understandings, agents would have no reasons to extend or defend their positions in a discursive interchange; no validity claims are redeemable between communicative participants if the agent cannot access, substantiate or understand their own rationality. The psychological and discursive preconditions that agents must manifest to meet Habermas's conditions as participants in communicative rationality are exceptionally demanding. This paper brings empirical research from psychology and political science, and conceptual critiques from political philosophy, to bear on Habermas's agent.*

The social psychology of communication is indebted to Jürgen Habermas. Many of the core tenets, derivative theories and critiques of the late 20th century concerning the social psychological phenomenon attendant to communication extend from his work. But as a discipline, social psychology has not been diligent in its ongoing evaluation of Habermas's core assumptions. While many critics are intent on attacking what is commonly seen as the central communicative tension of Habermas's communicative theory – the distinction between strategic and communicative action – other disciplines have taken significant strides in analysing the core assumptions of his programme. Work in political science and analytical philosophy, mainly in the domain of deliberative democracy, has given rise to questions about the processes and preconditions that Habermasian agents require to succeed in their deliberations. As one theorist attests, 'The similarities between communicative rationality and conditions of effective deliberation should be obvious. It is a straightforward matter to apply the components of communicative rationality as a set of criteria for deliberative democratic authenticity' (Dryzek, 2000, p.22). *Mutatis mutandis*, problems with one attend directly to the other. Deliberative democratic theory focuses not only on the optimistic possibilities that deliberation gives to democratic legitimation, but reconceptualizes through new eyes what it means to be a deliberative agent, and whether such agents might reasonably exist in modern, pluralist states.

The aim of the paper is to enter the agent problematic through a new gate – deliberative theory – before reviewing the Habermasian agent and attendant empirical psychology more comprehensively. We start by gathering recent conceptual problematics from deliberative democratic theory, particularly from the central authors in the field who are deeply versed in Habermasian themes. These authors are generally sympathetic to Habermas's aims, yet their political backgrounding forces them to bring agent limitations into the foreground. We argue that while deliberation exists, Habermas's construction of communicative rationality rests upon an agent role that might only be filled in reality by a

self-reflexive critical genius. Deliberative agents are assumed to be heroic in terms of informational breadth and calculative abilities, and heroic in their ability to identify, segregate and set-aside self-interest. This agent might be an individual of Habermasian proportions and Habermasian abilities, but they are no agent of modern actuality.

Although deliberative democracy critiques the agent and their dialogical requirements, there is a need to carefully review the agent as understood by Habermas. The second section of this paper surveys Habermas's agent requirements. The scrutiny of Habermas's construction of communicative rationality and the demands it makes on its agents gives a clear view of what Habermas expects in deliberation. To avoid criticisms of building a 'straw-man', the paper relies heavily on Habermas's primary texts. What the texts show is that the communicative rationality Habermas expects is not an ideal, but a necessary and pre-reflexive element of the lifeworld. With a clear picture of the agent, and an understanding of the conceptual problems that deliberative agents possess in the political realm, the agent must be subjected to psychological scrutiny. The third section of the paper submits the Habermasian agent to a brief empirical review of rationality, preference theory and deliberative practice. By abstracting from recent research in the domains of economic and political psychology, Habermas's agent is tested in the 'real' world. Empirical work in economic and social psychology pioneered by Amos Tversky and Daniel Kahneman is now asking deeper empirical questions about the actual operation of agents, and how far normative expectations of rationality might reasonably extend. These psychological research programmes have specifically questioned the normative assumptions that prevailing economic rationality models - like utility theory, consumer theory, and preference theory - are based on. The conclusion is a decisive one against Habermas. Social psychology must now test the empirical psychological outcomes of discursive rationalism in group environments *qua* Habermasian communicative rationality. This review suggests that the Habermasian agent is simply unavailable in the modern state. Furthermore, the procedures and institutions required to bring the Habermasian agent into being violate the key tenets of Habermas's own proceduralism. It is this Habermasian 'paradox' that the social psychology of communication must attend to.

## **1. Deliberative Democracy; new ways of reviewing communicative rationality and its participants**

When James Johnson notes, 'we might follow Habermas and argue that democratic deliberation embodies something like his concept of communicative reason' he introduces the link between Habermas's work on communicative action, communicative reason, and critical social theory, to the recent and highly productive work on deliberative democratic theory (Johnson, 1998, p.172). Deliberative democratic literature leans very heavily on the normative communicative frameworks that Habermas constructs. Deliberative democracy is a procedural political view that seeks democratic legitimacy through the capacity of those affected by a collective decision to deliberate in the production of that decision. Deliberation involves discussion in which individuals are amenable to scrutinizing and changing their preferences in the light of rational argumentation (not manipulation, deception or coercion) from other participants. Claims

for and against courses of action must be justified in terms participants can accept. Deliberative democrats believe that deliberation yields rational collective outcomes, in opposition to vote aggregation methods (social choice theory) which are susceptible to arbitrary and non-representative outcomes. Deliberative democracy is Habermasian communicative rationality writ large.

The notion that legitimate democracy revolves not around the mere collection, aggregation and reporting of private preferences, but depends upon the discussion and transformation of preferences, is now one of the major positions in democratic theory and has occurred largely ‘under the influence of Jürgen Habermas’ (Elster, 1998, p.1). John Dryzek, a leading deliberative scholar, agrees that by the late 1990s deliberative democracy provided the focal point for ‘much if not most democratic theory.’ John Rawls and Jürgen Habermas, ‘respectively the most important liberal theorist and critical theorist of the late twentieth century’, lent significant weight to the deliberative turn (Dryzek, 2000, p.1). Dryzek also recognizes that deliberative democracy’s theoretical roots lie substantially in critical theory, meaning that deliberative democracy must be grounded in ‘a strong critical theory of communicative action’ (Dryzek, 2000, p.4). It is the reopening of the deliberative debate on new battle lines that gives this paper an impetus to investigate the theoretical and empirical shortcomings of Habermas’s theory of communicative rationality. The most trenchant criticisms of Habermas’s theory concerning formal pragmatics and communicative rationality revolve around his distinction between communicative action and strategic action. These criticisms are valid, but misguided. Although some of the early critical literature reflecting on his communicative theories deals directly with his construction of communicative rationality (as we discuss in the following section), early critiques only touch briefly on the core of concern; the rationality *actually* available to the Habermasian agent. It is in the discovery and pursuit of this agent that new debates on deliberative democracy allow fresh analysis. Quite simply, Habermas is reliant on agents who, while explicable in theoretic terms, are practically unavailable in psychological terms.

Many of the most prolific and engaged writers of the deliberative democratic tradition have also been translators of Habermas’s work. James Bohman and William Rehg are at the forefront of deliberative debate, and acutely aware of Habermasian scholarship. James Johnson and Jon Elster have written extensively on Habermas and Habermasian themes, further strengthening the link between this field of political philosophy and Habermas’ work on communication and critical social theory. It is through these authors that we wish to conceptualize the problematic of the Habermasian agent, for the communicatively rational agent of Habermas is *mutatis mutandis*, the deliberative agent of deliberative democracy. It must be noted that all these authors are strong proponents of deliberative democracy, of public reasoning. They all acknowledge debts to Habermas’s scholarship and normative guidelines for deliberative interactions. Yet all of them recognize that deliberation *qua* communicative rationality exacts a significant if not impossible load on participating agents. Jon Elster, in discussing the similarities in Rawls’ notion of reflective equilibrium and Habermas’ ideal speech situation asserts that, ‘political choice, to be legitimate, must be the outcome of *deliberation about ends among free, equal, and rational agent*’ (Elster, 1998, p.5). The terms free, equal and rational are where the agent problematic might reasonably begin. If the transformation of preferences relies on successful communicative interchanges, and if

successful communicative interchanges rely on free, equal and rational agents, then we had best be able to point, in some meaningful empirical and substantive way, to the existence of such agents in the manner that satisfies Habermasian agency. Furthermore, the rational agents need to possess some remarkable qualities, not the least of which is an uncanny ability to scrutinize with great accuracy and humility, the complex relationships between a multitude of values and means-ends relationships. As Elster asserts:

‘arguing aims at the transformation of preferences ... much arguing is about factual matters. These statements are not inconsistent with each other. Individuals have fundamental preferences over ultimate ends and derived preferences over the best means to realize those ends, the gap between the two being filled by factual beliefs about ends – means relationships. Arguments that affect those beliefs will also affect the derived preferences’ (Elster, 1998, p.7).

It should be intuitively available to even the least critical observer that irreconcilable differences in perceptions of means-ends relationships are manifest in the world at large, they are also patently available in the psychological literature discussed below. If argumentation requires rational agents shifting positions under the weight of ‘the better argument’ with regards to particular means-ends relationships, Habermasian agents need to possess (universally), the cognitive predisposition to release certain preconceptions about exactly the means-ends relations they are contesting. That action is a necessary precondition of communicative reason. As Habermas (1996, p. 540) states ‘presuppositions of rational discourse have a steering effect on the course of the debates.’ And with its reliance on communicative reason, deliberative democracy incurs the same requirements: ‘Deliberative democracy rests on argumentation, not only in the sense that it proceeds by argument, but also in the sense that it must be justified by argument’ (Elster, 1998, p.9).

### *1.1 Are deliberative agents available?*

James Johnson outlines very clearly some hurdles that deliberative theorists need clear before deliberative democracy can claim any *prima facie* superiority over other mechanisms of democracy. In doing so, he explicates the deliberative core of the Habermasian agent:

‘[P]arties to democratic deliberation necessarily rely on communicative reason and, by raising and challenging the validity claim (to truth, normative rightness, and sincerity) that constitute the pragmatic presupposition of all linguistically mediated interaction, aim to establish a shared understanding of the context of their interaction ... this involves arriving at an agreement that is justifiable solely by reasons in light of generalizable interests of the relevant parties’ (Johnson, 1998, p. 173).

Johnson concludes that Habermas’s major stumbling block is the distinction between communicative and strategic action, but points to something more fundamental. Without a more completely theorized view of agents and agent expectations within Habermasian interchanges, ‘talk about “deliberative rationality,” “communicative democracy,” and so

on, remains significantly incomplete, as much a promissory note as a warranted commitment” (Johnson, 1998, p. 173). Johnson then poses the challenge to deliberative democracy. Without being overtly cynical about the realism of political agents, any defence of democratic deliberation must be reasonable in accommodating itself to current knowledge about political motivation and practical reasoning. In Johnson’s (1998, pp. 173-174) estimation it is unreasonable to ‘anticipate that deliberation will massively transform the preferences, capacities, or character of participants in normatively attractive ways’. To do so makes ‘heroic assumptions about participants.’

The same criticism must be levelled against Habermas, and in an even more fundamental form. Habermas’s communicative proceduralism is apparent in all his communicative theory since his ‘linguistic turn’ of the 1970’s. His discourse ethics, his communicative social theory, his juris theory, indeed his approach to all work since his communicative turn relies entirely upon heroic assumptions of agent rationality and agent deliberation that do not, in any reasonable sense, accommodate current conceptual or empirical knowledge about agent rationality, agent motivation, or the problems associated with consensus. The heroic reliance on agent reason, and public consensus via public reasons explored deliberatively, is simply untenable. To place it in a modern context; ‘A plausible argument for deliberation, in short, would acknowledge that substantive agreement on preferences or values is neither practically realistic nor normatively appealing in a large, pluralist constituency’ (Johnson, 1998, p. 176). Even Amy Gutman and Dennis Thompson (2004, p. 50), contemporary deliberative scholars of an optimistic orientation, accept that groups ‘intent on challenging the status-quo’ tend not to indulge in the ‘cool reason-giving’ that deliberative democracy requires. Agents are more likely to ‘use passion than reason. And for good reason: emotional rhetoric is often more effective than rational syllogism.’ As might be expected from Habermas, his political theory places deliberation at the core. Habermas intends that ‘the political system should transform preferences through public and rational discussion’, and that individuals do not consider their views in isolation, in selfish or purely self-regarding terms. Private wants, selfish wants are all exposed to the power of argument and universalization through public reasons and then, in Habermas’s view ‘uniquely determined rational desires would emerge’ (Elster, 1997, p. 11). The Habermasian explanation here, according to Elster, is that there is a ‘conceptual impossibility’ associated with expressing purely self-regarding arguments in a public forum. Indeed, in accord with what Elster (1997, p. 12) refers to as the ‘civilising force of hypocrisy’, there exists a psychological difficulty in ‘expressing other-regarding preferences without ultimately coming to acquire them’ such that public discussion should tend to promote the common good as individuals yield under the weight of universalizable public interests, and release their claims to pure self-interest. While Elster is a strong supporter of deliberative democracy he raises some very powerful conceptual objections that initiate reflection. Deliberative democracy, just like Habermas’s communicative reason, is in danger of being ‘dismissed as Utopian’ if it fails to grapple with the possibility that it might be ‘neglecting some elementary facts of human psychology’ (Elster, 1997, p. 13). While outlining Elster’s objections, we add that they are neither exhaustive nor conclusive – yet they conceptually preface the arguments we pursue empirically in the following section of the paper.

### *1.1a Informed preferences, informed participants*

It is difficult to intuitively assign a similar value to the inputs and rights of a non-informed and/or uninterested participant in a communicative interaction. Moreover, what must an agent's situation be such that they have the resources and skills to pursue both information *and* involvement? Surely, as Elster asks, 'one should try to favour informed preferences as inputs'? And might this not possibly mean that the body of participants who can leap the hurdle for participation are 'disproportionately found in the privileged part of the population'? The outcome is that 'the high ideals of rational discussion could create a self-elected elite' (Elster, 1997, p. 14). Informed agents, who are motivated to engage, reflect and accurately weight the respective options under consideration do not seem universally available in the modern public sphere. Where they are available, they are available selectively.

### *1.1b The problem of plurality*

The next heroic requirement of deliberative / Habermasian agents is their requirement to abandon personal perspectives if somebody can provide a better argument that *their* view of the good life, is not *the* view of the good life. The central human values that inform an individual's life programmes are notoriously difficult to reengineer. Thus, 'even assuming unlimited time for discussion, unanimous and rational agreement might not necessarily ensue.' Might we then conclude that there might exist 'legitimate and unresolvable differences of opinions over the nature of the common good?' (Elster, 1998, p. 9). Would Habermas deny the possibility of ultimately irreconcilable pluralism in values, and irreconcilable pluralism between the agents who maintain these values? As McCarthy asks, 'If the variety of worldviews and forms of life entails an irreducible plurality of standards of rationality' then surely 'the concept of communicative rationality could not claim universal significance and a theory of society constructed on it would be limited from the start to a particular perspective' (McCarthy, 1984, p. xi).

### *1.1c Some agents learn faster, or learn more, than others*

As Elster (1997, p.15) points out, for some people, no discussion is better than some discussion: 'an inferior outcome may result if discussion brings about partial adherence to morality in all participants rather than full adherence in some and none in others.' Again this begs the question of what powers the deliberative agent is assigned in Habermasian theories of interaction. Given the same inputs, and the same levels of informational determinateness, will all agents reach the same conclusions? Might partially informed agents, making assumptions about the 'wider programme' of good that they have imputed from only partial understandings or fragmented rationality, end up doing more damage than good? More significantly, what impact will they have on the wider programme of public reasoning?

### *1.1d Is a whole smarter than the sum of its parts anyway*

Elster (1998, p. 9) questions ‘the implicit assumption that the body politic as a whole is better or wiser than the sum of its parts’. The cognitive analogy begs the question whether the ‘rationality of beliefs may be positively as well as negatively affected by interaction’. ‘Could it not rather be the case that people are made more, not less, selfish and irrational by interacting politically?’ A choice to engage in deliberation does not apply a compelling requirement for immediate rationality to all participants, and there is nothing to impute from deliberation that suggests ‘the massive and coordinated errors that may arise through group-think’ might necessarily be avoided. The avoidance of such errors relies again on the heroic assumptions concerning communicative participants.

### *1.1e Unanimity is not accuracy*

The last major objection of Elster’s we wish to canvass is the very real psychological fact that unanimity, the desired Habermasian outcome, might be as easily achieved through conformity as rational agreement.<sup>i</sup> Again, there seems no necessary preconditions for deliberation that ensure agreements reflect *desirable* cognitive mechanisms, or desirable outcomes. As Elster (1997, p. 15) voices, ‘I would in fact tend to have more confidence in the outcome of a democratic decision if there was a minority that voted against it, than if it was unanimous... Social psychology has amply shown the strength of this bandwagon effect.’ Habermas (1993, p. 441) himself adds weight to this when he asserts: ‘Tocqueville and John Stuart Mill were perhaps not so mistaken in their belief that the early liberal notion of a discursively accomplished formation of opinion and will was nothing but a veiled version of majority power’. The tyranny of the majority may not be the best surrogate for the rational pursuit of the good life.

James Bohman extends this theme of questioning the social psychological and cognitive validity for the strong agent assumptions underwriting deliberative success. In an article resting heavily on Amartya Sen’s review of human functioning and human flourishing as products of equal capability sets, Bohman analyses the agent problematic in a slightly different light. He concurs that deliberative ideals place ‘great demands upon citizens’ abilities and willingness to express their own reasons publicly and consider the public reasons of others.’ From this extends an equally ‘demanding ideal of equality’ (Bohman, 1997, p. 322). Bohman highlights a demanding notion of equality, yet also highlights equally demanding notions of capability - capability here in the sense of heroic agency, disinterested universalisation, and agent-neutral public reason. Habermas’s proceduralist communicative rationality requires deliberations that pass the test of publicity. Yet the procedures are, psychologically speaking, neither necessary nor sufficient for guaranteeing compellingly ‘good’ public reasons. For Habermasian communicative action, like Bohman’s critique of deliberative democracy, ‘proceduralist accounts’ are ‘guided by inadequate and incomplete conception of political equality, namely, equality of opportunity’ (Bohman, 1997, p. 323). One might want to retort that Habermas has very strict procedures about equality of power and absence of threat in his construction of conditions for communicative action, but he says very little about the qualities required of his agents, of the socio-political reality that actually meet his heroic requirements. As Bohman (1997, p. 326) laments:



Employing public reason in dialogue with others clearly requires highly developed capacities and skills related to cognition and communication...if deliberative politics is to remain democratic, it cannot simply favour those who are most educated, who have access to special information, who possess the greatest resources and privileged social positions'

Bohman goes on to question the equality of capabilities available to the agents from which successful deliberation extends. The success of deliberation 'is dependent on convincing others about the cogency of one's reasons and on judging the cogency of the contributions of others.' To effectively deliberate, 'citizens must still know something about how to deliberate, how their reasons will be responded to, and how their goals may be achieved. They must know what it means to succeed in deliberation' (Bohman, 1997, p. 337). And so what minimum demands must be met for our Habermasian agents, or our deliberative democratic agents under Bohman's view, to be successful? Beyond merely possessing information, they must be able to analyze and synthesize it, converting it into convincing public reasons. They must also have access to the appropriate forms of communication that might make deliberative success a possibility, and they must also have the capability to use informational and personal resources to deliberate with others in public (Bohman, 1997, p. 342). The problem here is that the lifeworld of such agents, with their equal access and use of relevant information, requires something seemingly unavailable in the modern socio-political context. It requires heightened agent understanding of reflexivity and publicity and universal adherence to these norms. It requires recognition and suspension of private reasons in a world dominated by self-interest. It requires the suspension of that same self-interest which economic and policy spheres are seemingly built upon. Classical and neo-classical economics, and all the derivative policies extending from it assumes instrumental rationalism and self-regarding actions. Modernity, with its increasing access to information and technology is not making Habermasian agents more likely, nor is it increasing the chances of deliberative success.

## **2. Reviewing Habermas's agent rationality and communicative rationality**

After using deliberative theorists to open the agent problematic, it is time to now comprehensively explore the Habermasian agent. Habermasian concepts lend themselves easily to deliberate misrepresentation. Habermasian themes can be unfairly abstracted and then brutalized with glee; an accusation Habermas (1982, p. 263) even directs at Anthony Giddens. Criticism of Habermas has found its mark in certain cases, and on occasion he has rigorously reformulated his perspectives, or strengthened his theoretical or conceptual enunciations. His work on communicative action has drawn a strong and sustained critique, particularly in the decade after its original release and subsequent English translation. We survey some of these criticisms and while the insights are telling, they still fail to attack the Habermasian agent in the manner which the recent work in deliberative theory invites. For those expecting to dismiss this component of our argument with the standard Habermasian defence of 'idealized states' and 'ideal speech situations', the defence is a difficult one to mount. Although Habermas includes elements

of idealization in his theories, communicative rationality is the necessary assumption that underwrites communicative action, not an idealization.

### 2.1 Habermas's agent rationality

The boldest expression of Habermas's expectation for his communicative agents comes when he states:

If we assume that the human species maintains itself through the socially coordinated activities of its members and that this coordination has to be established through communication – and in certain central spheres through communication aimed at reaching agreement – then the reproduction of the species also requires satisfying the conditions of rationality that is inherent in communicative action (Habermas, 1984, p. 397).

Habermas (1984, p. 398) goes on to assert that life as a 'process of self-preservation' has to 'satisfy the rationality conditions of communicative action' because subjects integrate their achievements and accomplishments through action coordinated via criticizable validity claims. These are strong declarations, and must be heeded by those who want to apply the 'idealized' defence to Habermas's rationality constructs. Habermas is stating that if we assume human activity is maintained through socially coordinating communication, then the fact the human species persists, extends from rationality assumptions implicit in human interaction, as McCarthy (1982, p. 66) adds:

Habermas has to show that the ability to act communicatively (in his strong sense) and to reason argumentatively and reflectively about disputed validity claims is a developmental-logically advanced stage of species-wide competences, the realization and completion of potentialities that are universal to mankind. To anyone familiar with the 'rationality debates' that have accompanied the development of cultural anthropology from the start...it will be clear that the burden of proof is considerable.

Not only is the burden of proof considerable, there is also nothing idealized in Habermas's claim. Habermas is stating that his vision of communicative rationality is developmentally and logically necessary. Whether he then successfully constructs communicative action as an idealization is immaterial to this paper; we are not concerned with communicative action *per se*. What we are concerned with is that the quotes above commit communicative rationality as necessary and foundational in Habermas's programme. For communicative rationality to be foundational, Habermas needs rational agents. Habermas (1984, p. 106) himself states; 'Communicative actions always require an interpretation that is rational in approach'; that 'the expression "rational"' is 'a guideline in elucidating conditions of rationality both for speaking and acting subjects and for their expressions.' What that rational agent might be is stipulated by Habermas in no vague manner.

Perhaps what is most startling about the rationality elucidations in the opening chapters of *The Theory of Communicative Action* (is that Habermas commences with what might be considered an unassuming and highly accessible conceptualization of

rationality. Habermas (1984, p. 8) seems not to depart radically from the intuitive meanings of individual rationality, or from classical conceptions of the rational agent.

What does it mean to say that ... expressions can count as “rational”? ... The close relation between knowledge and rationality suggests that the rationality of an expression depends on the reliability of the knowledge embodied in it.

Certainly Habermas outlines a rationality that is dependent on subjective interpretations, but his denial of relativism commits him to subjects and objects that admit to (eventually) *factuality* in the strongest sense of the word. He states clearly that the ‘rationality’ of ‘expressions is assessed in light of the internal relations between the semantic content of these expressions, their conditions of validity, and the reasons (which could be provided if necessary) for the truth of statements’. The rationality of an expression extends from its ‘being susceptible of criticism and grounding: An expression satisfies the precondition for rationality if and insofar as it embodies fallible knowledge and therewith has a relation to the objective world (that is, a relation to the facts) and is open to objective judgement’ (Habermas, 1984, p. 9). One might even argue there is something distinctively positivistic about Habermas’s rationality, that it has a ring of communicative verificationism to it; ‘assertions and goal-directed actions are the more rational the better the claim (to propositional truth or to efficiency) that is connected with them can be defended against criticism’ (Habermas, 1984, p. 9).

While Habermas (1984, p. 66) might literally reject a requirement for instrumental rationality in his agents, arguing that the West’s fixation on cognitive-instrumental rationality is ‘a distorted understanding of rationality’ requiring a shift from ‘cognitive-instrumental rationality to communicative rationality’, the majority of assumptions attendant to instrumental rationality are still implicitly required by his communicative theory. Despite his assertion that what is central to his agents is intersubjective agreements, and not ‘the relation of a solitary subject to something in the objective world that can be represented and manipulated’, those agents still need to possess the substantive traits of instrumental rationality, (for example, classically goal rational individuals) if they are to meet the rest of Habermas’s preconditions (Habermas, 1984, p. 392). As stipulated by Habermas (1984, p. 21), those rational preconditions demand an agent who:

- a) ‘is capable of letting himself be enlightened about his irrationality’,
- b) is in possession of faculties that allows him to ‘judge facts’,
- c) acts in ‘a purposive-rational way,’
- d) is ‘morally judicious and practically reliable’,
- e) will evaluate matters ‘with sensitivity’ and,
- f) will possess ‘the power to behave reflectively in relation to his subjectivity and to see through the irrational limitations to which his cognitive, moral-practical, and aesthetic practical expressions are subject.’

These seem stringent conditions for agency, and they bear many of the hallmarks of classic agent rationality, including Habermas’s (1991, p. 229) requirement that his agents manifest objective preference orderings. To further clarify the demands that weigh on the

Habermasian agent, several additions to the list of agent preconditions are required. Again according to Habermas (1984, pp. 19-22, p. 134), the Habermasian deliberative agent would:

- g) not enter into ‘moral argumentation if he did not start from the strong presupposition that a grounded consensus could in principle be achieved’
- h) possess the ability to construe and test norms where ‘valid norms must be capable in principle of meeting with the rationally motivated approval of everyone affected under conditions that neutralize all motives except that of cooperatively seeking the truth.’
- i) ‘speak and act in modes of behaviour for which there are good reasons or grounds.’
- j) make ‘rational expressions [that] admit of objective evaluation’
- k) acquire through learning processes ‘theoretical knowledge and moral insight, extend and renew our evaluative language, and overcome self-deceptions and difficulties in comprehension’
- l) be ‘oriented to reaching understanding and thereby to universal validity claims’
- m) base ‘their interpretative accomplishments on an intersubjectively valid reference system of worlds’
- n) possess ‘a decentered understanding of the world.’<sup>ii</sup>

These are the necessary traits that individuals must carry into all their communicative interchanges (the heroic requirements mentioned at the beginning of the paper). Habermas (1991, p. 255) explicitly acknowledges the stringency of his demands:

My position is that those who understand themselves as taking part in argumentation mutually suppose, on the basis of the pre-theoretical knowledge of their communicative competence, that the actual speech situation fulfils certain, in fact quite demanding, preconditions.

The acknowledgement however does nothing to lighten the heroic load. Habermas stipulates how the agent is to be constructed, and the stipulations should worry communicative theorist and psychologist alike.

## *2.2 Habermas’s Communicative Rationality*

Habermas’s claims for agent rationality move from the initially unassuming, through to the increasingly unlikely. Yet his expectations of communicative rationality are immediately inflationary:

communicative rationality carries with it connotation based ultimately on the central experience of the unconstrained, unifying, consensus-bringing force of argumentative speech, in which different participants overcome their merely subjective views and, owing to the mutuality of rationally motivated conviction,

assure themselves of both the unity of the objective world and the intersubjectivity of their lifeworld (Habermas, 1984, p. 10).

Habermas (1984, p. 14) also asserts that only 'responsible persons can behave rationally' and in the context of communicative action, 'only those persons count as responsible who, as members of a communication community, can orient their actions to intersubjectively recognized validity claims.' Yet if two agents possess convictions which both believe are rationally motivated, but cannot be subjectively reconciled nor objectively proven, how are the agents to become communicatively unified? If an agent maintains community delusions merely because it is the norm to do so, how are they motivated to seek objective verification of those collective delusions?

Some of the most perspicacious assertions concerning communicative rationality come in Habermas's essay "A Reply" in Honneth and Joas's edited collection *Communicative Action*. Habermas quickly, and directly, locates communicative rationality not as an idealization, but as a concrete and necessary form of social understanding. Recognizing his debt to Humboldt, Habermas (1991, p. 220) explains that Humboldt had already given a 'normative twist' to formal pragmatics, therein supporting the 'rational potential of speech' in what is admittedly 'the necessarily idealizing suppositions of communicative action'. Many Habermasian critics approach the problematic of communicative action through its heavy reliance on 'idealizing suppositions'. Yet fewer attack the non-idealized assumptions of communicative rationality. It is exactly these assumptions of agent communicative rationality that Habermas needs as a precondition to even consider communicative action as a normative framework.

Communicative rationality is not merely submitted by Habermas as normatively valid, he goes further to claim that one might 'reasonably assume' that communicative rationality 'can be used in a purely descriptive manner'. He suggests we might 'impute to communicative actors that they cannot avoid making certain pragmatic assumptions, harbouring idealizations and orienting themselves towards validity claims'. This analysis asserts that agents 'are subject to normatively substantive rationality constraints' and in this sense 'submit to an objective order of preferences' (Habermas, 1991, p. 214). Two points must be contested here. Firstly, that Habermas claims his theory holds as descriptively valid is bold; normative rationality models are under attack by recent psychological research. Moreover, a *descriptive* model with the stringent agent requirements that Habermas outlines, is even less justifiable than a normative one (as we argue in more detail in the following section). Secondly, that agents submit to objective preference orderings in any meaningful sense is another contestable claim - particularly so given the amount of theoretical and empirical work that attacks preference theory (again, as detailed in the following section).

Preference theory has extensive problems, not the least of which is the assumptions it requires concerning agents' access to the reasons for the preferences they hold. The problem of preferences, preference ordering, and privileged access to preferences is an active area of social science research, and most recently the greatest thorn in the side of economic theory. 'If we can imagine an individual making a complete and transitive ordering over possible outcomes over time, then inter-temporal comparisons are implicit in the preference orderings and cause no particular difficulty

beyond the heroic character of the assumption about human capabilities' (March, 1986, p. 159). As McCarthy adds, Habermas is 'after a notion of ego identity that centres around the ability to realize oneself under conditions of communicatively shared intersubjectivity.' Habermas (1984, p. xxi) demands his agents 'act in a self-critical attitude.' It is the universal 'expectation that the participants question and transcend whatever their initial preferences may have been' that makes the agents so entirely heroic (Habermas, 1993, p. 449).

### *2.3 Who are these Habermasian agents?*

In some of the earliest and only agent-related criticism of communicative rationality, Steven Lukes attends to the question of how communicative action attains its sublimely rational status. The answer - it is populated with sublimely rational agents:

If we are asked to imagine what ideally rationally agents would do under the posited conditions, the whole argument turns on the nature of those agents and the constraints set by the conditions. If these together are such that the appropriate answers are necessarily reached, then the counterfactual hypothesis emerges as vindicated but only because it has been so formulated that it must do so. Ideally rational people in an ideal speech situation cannot but reach a rational consensus (Lukes, 1982, p. 140).

Lukes (1982, p. 141) echoes the ongoing concern of this paper when he asks: 'Who are the participants in the unconstrained discourse that is held to offer the possibility of rational consensus.' Where might we find these individuals? The rationality prescribed by Habermas initially is not entirely unreasonable. Yet the cumulative conditions that must obtain for agents universally in their dialogical representations is where the unreasonable stringency is introduced; 'under the imagined conditions of ideal communication, actual actors would be so transformed as to become capable of the requisite rational consensus' (Habermas, 1991, p. 221). As Lukes (1982, p. 141) addends, this is no small ask; 'there is every reason to suppose that this would not be so, if socialization processes and relations of economic and political power remain unchanged'. With respect to Habermas's attribution of objective preferences to his agents, the assumption is that communicative rationality will lead to a review, reassessment and rearrangement of these preferences in all cases, as that is the necessary and sufficient condition for consensus; the idea that 'there will be an endogenous change of preferences on the part of social actors.' Lukes (1982, p. 145) adds that the universalizability of judgements 'does not impose rational constraint on choices of action or defensible patterns of behaviour'; at any point it is both logically possible and common for people to avoid those very rational constraints.

### *2.4 Is there humanity in Habermasian agents?*

Agnes Heller, in her piece questioning Habermas's communicative rationality, wonders precisely where Habermas leaves space in his theories for the more mundane, and less efficient aspects of human living, the aspects that make humans so very real. She laments that Habermas leaves no room for 'sensuous experiences of hope and despair, of venture and humiliation' accusing him of completely avoiding 'the creature-like aspects of human beings' (Heller, 1982, p. 21). Heller (1982, p. 21) points out that although

Habermas does outline a differentiation between merely living and an Aristotelian notion of a “good life”, it seems difficult to resist the conclusion that for Habermas, the good life is a life marked by endless rounds of rational, decentred, agent-neutral communication – where human needs are argued for ‘without being felt.’ Heller (1982, p. 25) goes so far as to charge Habermas with disregarding the ‘whole motivational structure of human beings.’

If we accept the plurality of ways of life, we have to accept the plurality of theories as well. Consensus regarding one theory would mean consensus in one single way of life. To exchange pluralism for consensus would be a bad bargain (not only for me, but for Habermas as well). (Heller, 1982, p. 31).

Heller pursues Habermas’s claims that progress ‘can be nothing but the realization of rationality, with priority given to communicative rationality’. In this way, Habermas allows only one path forward; progress as rationality. Heller (1982, p. 43) coaches like many others who resist Habermas’s proceduralism; that procedure does not substantiate outcome - ‘domination-free argumentation can be conceived only as a precondition of the good life – it is not the good life itself.’

Habermas replies to Heller that his thesis for rationality is comparatively conventional; a defence supportable only to limited degrees. It is in his assumption that individual rationality must, and will, manifest itself within communicative rationality, where he steps too far. Habermas’s (1982, p. 277) programme would have us accept that agents ‘cannot avoid having recourse, intuitively, to foundations that can be explained with the help of the concept of communicative rationality’ and that in communicative rationality agents are always oriented toward the validity claims upon which possible consensus depends. Habermas (1982, p. 228) even asserts that ‘[c]ommunicative reason operates in history as an avenging force.’

### *2.5 The Habermasian Agent - Conclusion*

According to Habermas himself, the ‘concept of communicative rationality does not just apply to the processes of intentional consensus formation, but also to the structures of a state of pre-understanding already reached within an intersubjectively shared lifeworld.’ This lifeworld, Habermas’s ‘context-forming horizon’, is not a state of idealization. This lifeworld is ‘an unproblematic and prereflexive background’ that ‘plays a constitutive role in the achievements directed toward reaching understanding’ (Habermas, 1991, p. 223). This unproblematic background is where Habermas’s communicative rationality lives, not in his idealizations. For this reason communication theorists and psychologists alike should scrutinize the empirical and conceptual validity of Habermas’s claims. The idealizations in his theory slip easily from grasp, but unproblematic and prereflexive background should be fair grounds for robust testing, and discovery, on the true nature of rationality in a communicative domain.

### **3. The search for Habermas’s agents**

The sheer volume of findings in such areas as heuristics, biases, decision theory, cognitive dissonance, prejudice, ingroup bias, nonconscious process, and a host of others imposes on an intelligent outside observer the view that the gap between the ordinary person's everyday judgement and behaviour and what a normative rational model would require is vast. ... The rationalist position has only scholarly merit (Zajonc, 1999, p. 211).

Thomas McCarthy states that the 'structures of communicative rationality' as constituted by Habermas 'would have to prove themselves adequate to the empirical materials' in areas such as psychology and anthropology. The proof is required because the universal significance Habermas requires 'cannot be established solely by the 'horizontal' reconstructions of formal pragmatics, it would have to be established principally in the 'vertical' reconstructions of developmental theories'. For McCarthy (1982, p. 68), and in terms of the argument we are building here, 'the mastery of these structures would have to be shown to represent the developmental-logically most advanced stage of species-wide competencies.' What is more, Habermas would need to show that the mastery of such structures applies universally. When McCarthy (1982, p. 72) advises that Habermas would 'be well advised to adopt a much more tentative and critical posture toward cognitive developmental theories than he has to date' for the reason that they 'cannot be appealed to as providing confirmation of his universality claims', he accurately envisages the empirical research programmes and conceptual advances that make Habermas's assumptions about preconditions of communicative rationality so difficult to uphold. That Habermas (1993, p. 439) has actually acknowledged as much, claiming, 'I cannot even begin to comment on the diversified literature in the sociology of political behaviour, since I have only paid sporadic attention to it', seems insufficient explanation for somebody offering a communicative theory that claims both normative and descriptive validity. Dryzek (2000, p. 26) levels the same charge as McCarthy, asserting that Habermas's communicative rationality founders because it does not substantively connect with any 'empirical realities that are its alleged justification', arguing that 'Habermas falls prey to a familiar, indeed near-universal, tendency among political theorists: to treat empirical reality in terms of a few stylized facts, rather than attending seriously to the findings of empirical political science.'

So what is the evidence concerning human rationality in a communicative context? Much of the vigorous research in this respect has been carried out in political philosophy and economic psychology. Even here however, the conclusions are still highly illustrative, particularly given the breadth of the impact of rationality assumptions in the social science arena. Indeed, the implicit assumption of rationality in the social sciences seems almost ubiquitous. As Kahneman (2000, p. 758) attests it is 'central to much theory in the social sciences.' Further: 'it is commonly assumed that most if not all economic and political agents obey the maxims of consistency and coherence leading to the maximization of utility' (Quattrone and Tversky, 2000, p. 451). However, as recent research demonstrates, there is 'compelling evidence that the maintenance of coherent beliefs and preferences is too demanding a task for limited minds' (Kahneman, 2000, p. 774). Limited minds are exactly what human agents possess. The empirical findings are admittedly only recent, and no doubt subject to improvement, but this is precisely why this paper draws attention to the foundations of Habermas's communicative theory.



In a historical survey of experimental social psychology, Zajonc points to a longstanding schism in the discipline concerning the rationality of human behaviour: 'difference between the one side and the other depends on how much control over behaviour is accorded to voluntary and wilful reason and how much to the forces of nature and biological dispositions'. The two extremes are marked thus: at one end the agent is 'capable of adapting to and subduing the environment by the power of the intellect' and at the other end the agent is subject to 'the interplay of elementary processes, such as conditioning, discrimination, generalization, drives, habits' (Zajonc, 1999, p. 202). Zajonc (1999, p. 203) agrees that rationality has always lurked as an implicit assumption in social psychological theories and the theories of sibling disciplines. Most of the attention to testing those assumptions has fallen to economic psychology, courtesy of Tversky and Kahneman. Zajonc's survey is comprehensive, selecting a series of experiments and theories that purport to show rationality as either prevalent or lacking. Lorge (1936), he notes, demonstrated 'prestige suggestion' as a manifestation of irrationality. The early work of Festinger (1950) 'definitely accepts a rational view of the person' even though the cognitive dissonance processes highlighted demonstrate that the manner in which inconsistencies is solved is rather irrational (Zajonc, 1999, pp. 206-09). Although Zajonc's piece is thorough in its coverage of the rationality question in social psychology, it also highlights the recent dearth of active research in the area of rationality, specifically by social psychologists. Without labouring this point further, the Zajonc link demonstrates that rationality is a central platform of concern for social psychology, and social psychology itself holds an acknowledged debt to the work of Tversky and Kahneman for empirically testing the normative assumptions central to models of normative rationality.

### *3.1 Classical rationality and rationalization: what is the difference?*

Shafir, Simonson and Tversky highlight the relationship between reasons-based-choice and rational choice. They explain that the classical rationality model used in economics, decision theory and management science typically associates a numerical value with various alternatives, and rational choices are maximization choices. The alternative tradition, 'characteristic of scholarship in history and the law, and typical of political and business discourse, employs an informal, reason-based analysis.' In the opinion of the authors, little contact has been made between the two arenas, but they are not incompatible: 'reason-based accounts may often be translated into formal models, and formal analyses can generally be paraphrased as reason-based accounts' (Shafir, Simonson and Tversky, 2000, p. 598). Reasons-based accounts however, are not the Reason that Habermas wants, because to take this route creates the problem of rationalization versus rational. Rationality and rationalizations must be treated differently. 'An explanation of choice based on reasons ... is essentially qualitative in nature and typically vague. Furthermore, almost anything can be counted as a "reason," so that every decision may be rationalized after the fact' (Shafir, Simonson and Tversky, 2000, p. 619). When Habermas asks for public reasons, he is not talking about reasons after the fact to explain why something was done. What Habermas's 'force of the better argument' refers to is reflexive preconsideration; the why something ought be done. These are not rationalizations, this is rationality. Furthermore, it is rationality in a manner

that befits the classical model of rationality. Ex ante explanations are not what Habermas expects of his agents, which is understandable, because as has been ‘amply documented’, subjects do not even know what drove their choices, and generate spurious explanations when asked to account for their decisions (Shafir, Simonson and Tversky, 2000, p. 599).

Even though classical theory explorations of agent rationality are used here to review Habermasian agency, it is important to note that a purely ‘reasons-based’ approach offers little respite. Reason-based rationality might ‘accommodate framing effects ... and elicitation effects’ which explain how preferences are moulded by the ways in which options are presented, but it requires people to not have well-established values or preferences, requiring merely that their preferences be constructed – not merely revealed – during their elicitation (Shafir, Simonson and Tversky, 2000, p. 618). How might agents be expected to debate, defend, or explain their preferences and values if these apparently substantive beliefs were not available until such point as preferences were publicly expressed? Although the reason-based rationality theory may seem intuitively attractive to the Habermasian ideal, and might make descriptive models of rationality all the more permissive, it does not attend to the primary requirements that Habermas requests of his agents. Giving reasons after the act is not what Habermas seeks of his agents; giving the selfless and compelling reasons why one ought to act is what he demands.

### *3.3 An empirical psychological problematic for Habermasian agents*

In Pronin, Puccio and Ross’s examination of misunderstanding, the social psychological aspects of interpersonal and intergroup communication are reviewed with respect to rationality. The authors conclude that ‘People and groups who disagree about matters of mutual concern not only interact in conflictual ways; they also interpret, and frequently misinterpret, each other’s words and deeds’ (Pronin, Puccio and Ross, 2002, p. 636). Moreover, as suggested above, ‘the recipients of persuasive arguments often prove to be rationalizing rather than rational agents, and as such are influenced less by logical rigor or objective evidence than by the interests and preconceptions that they bring to their task’ (Pronin, Puccio and Ross, 2002, p. 636). During deliberation, during communicative interactions, participants might assume they are being rational, but empirical evidence points very much to processes of the mind demonstrating that they are in fact merely *rationalizing*, explaining away their newfound beliefs, or explaining away their refusal to drop their old beliefs. Rationalizing agents, persons who bring reasons to explain what they do, are not the agents that Habermas wants: he needs agents to give public reasons that are rational in Habermasian sense. Furthermore, the agent needs to be reflexive enough to understand the difference between public reasons for ‘the good’, versus reasons given publicly to defend any good. What’s more, there are psychological effects in play that encourage agents to understate other people’s rationality, and overstate their own. The study highlights the difficulty of consensus: ‘blindness about the role that ... biases play in shaping our own political views, and a penchant for seeing self-serving or ideologically determined biases in other’s views ... exacerbates group conflict’ (Pronin, Puccio and Ross, 2002, p. 637).

Using Piaget as a point of departure, the authors explain that people do indeed develop ‘skill at anticipating specific sources of perceptual, cognitive and motivational

bias' but what rationalists tend to then overlook is that in Piaget's account the 'process never reaches fruition... adults continue to show important limitations in perspective taking' (Pronin, Puccio and Ross, 2002, pp. 641-642). The 'lay epistemology' that limited perspective-taking results in is termed naïve realism and leads people to assume that their world view enjoys 'particular authenticity' and that other agents will share exactly that worldview 'if they are attentive, rational, and objective perceivers of reality and open-minded seekers of the truth.' As a result, misunderstandings and disagreement grow from agents' 'failure to recognize the operation of such biases in their own judgements and decisions'. Agents then even more firmly clinging to 'misguided notion' that that they themselves 'see the world, and evaluate divisive issues, in a uniquely clear, unbiased, and "unmediated" fashion' (Pronin, Puccio and Ross, 2002, p. 641). The end result pushes Habermasian agency further from empirical reality:

Cognitive biases [lead agents] to see and remember a reality that is consistent with their beliefs and expectations, while motivational biases cause them to see what is consistent with their needs, wishes and self-interest. Through such information-processing biases, two opposing partisans who encounter the same facts, historical accounts, scientific evidence, or even witness the same events can find additional support for their preconceptions (Pronin, Puccio and Ross, 2002, p. 649).

How does an agent, who must meet the agency stipulations spelled out above counter the evidence offered here? How does a Habermasian agent fulfill their duty and at the same time acknowledge the cognitive impossibility of achieving the 'view from nowhere'? Further evidence pointing towards the limitations of the agent is not difficult to compile. Tversky and Kahneman (2000, p. 210) argue that 'the logic of choice does not provide an adequate foundation for a descriptive theory of decision making ... deviations of actual behaviour from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the normative system.' Here it is important to point out that two of the four substantive assumptions associated with utility theory *qua* individual rationality are dominance and invariance. Dominance is an obvious principle of rational choice, and one relevant to Habermasian communicative reason. It simply states that if one option is better than another in one state and at least as good in all other states, the dominant option should be chosen. Dominance 'serves as the cornerstone of the normative theory of choice' (Tversky and Kahneman, 2000, p. 211). The next essential condition for a theory of choice is invariance, 'which states that the relation of preference should not depend on the description of the options (description invariance) or on the method of elicitation (procedure invariance)' (Slovic, 2000, p. 490). Tversky and Kahneman conducted a series of experiments from which they concluded that failures of invariance and dominance were explicable with reference to the framing of prospects, the evaluation of outcomes and the weighting of probabilities. Essentially their series of experiments found that people's preferences, their rationality, was different in outcome to what the normative theory of choice predicted. In judgements of probability, or visual assessments, how the problem was framed impacted greatly on how experimental participants responded in the ranking of their preferences. The significance to communication

derivative theories cannot be understated in these empirical reports; particularly the framing and certainty effects. As Tversky and Kahneman (2000, p. 211) argue, ‘the framing of decisions depends on the language of presentation, on the context of choice, and on the nature of the display’. As questioned by the deliberative theorists discussed above: how does an agent know if they are deliberatively competent? How does an agent know if the agreement being elicited from them is based on the fundamentals of a better argument, or merely a better arguer?

### *3.4 An empirical deliberative problematic for Habermasian agents*

Another empirical problem Habermas has extends from what Sunstein (2003, p. 81) refers to as the law of group polarization, the phenomenon that ‘members of a deliberating group predictably move toward a more extreme point of view in the direction indicated by the members’ predeliberation tendencies.’ It is a phenomenon that Sunstein (2003, p. 82), in an investigation pointed toward deliberative democracy, refers to as a ‘striking but largely neglected statistical regularity’. It is also ‘among the most robust patterns found in deliberating bodies, and it has been found all over the world and in many diverse tasks.’ Pronin also mentions this tendency, explaining that when ‘people earnestly attend to the facts and arguments offered by those on the “other side,” their opinions became even more polarized. Polarization, he notes, ‘reflects the tendency for partisans to accept at face value arguments and evidence congruent with their interests and beliefs, while critically scrutinizing arguments and evidence that threaten those interests and beliefs’ (Pronin, Puccio and Ross, 2002, p. 637).

Two principal mechanisms underlie group polarization. The first relates to social influences on behaviour and agents’ desires to retain reputations and self-perceptions. The second relates to the limited argument pools within groups. This is exceptionally problematic for deliberation, and therefore for communicative rationality generally, particularly from the normative perspective. ‘If deliberation predictably pushed groups toward a more extreme point in the direction of their original tendency, whatever it may be, do we have any reason to think that deliberation is producing improvements?’ (Sunstein, 2003, p. 82). The polarization sees a shift, ‘not toward the middle of the antecedent dispositions, but toward a more extreme position in the direction indicated by those dispositions.’ The effect is to both decrease variance, and to produce convergence on a ‘relatively more extreme point among predeliberation judgements’ (Sunstein, 2003, p. 83). Habermasian agents are certainly required to decrease variance, but not through convergence upon an extreme perspective. Indeed, increased homogeneity within the group tends ‘to suppress dissent’ and ‘lead to inferior decisions’ (Sunstein, 2003, p. 85). The normative preconditions that Sunstein suggests be instituted to avoid the problems of group polarization are demanding: there needs to be social spaces for deliberation by like-minded persons, but differences need to be sought also. Conversational goals must promote the interests of groups inside and outside the relevant enclaves, and views must be exchanged in a group-neutral manners ensuring that the wider society ‘does not marginalize, and thus insulate itself from, views that may turn out to be right, or at least informative’ (Sunstein, 2003, p. 91). Although the institutional demands are extensive, so too are the agent demands implicit in such institutions. Given the empirical evidence that ‘people are shifting their position to maintain their reputation and self-conception’

why might Habermas conclude that ‘deliberation is making things better rather than worse?’ (Sunstein, 2003, p. 91). Sunstein (2003, p. 92) drives the point home:

Perhaps group polarization could be reduced or even eliminated if we emphasized that good deliberation has full information as a precondition...But this precondition is extremely stringent...In any case the group polarization phenomenon suggests that in real-world situations, deliberation is hardly guaranteed to increase the likelihood of arriving at truth.

Another set of related empirical data for communicative rationality is presented by Kahneman and Tversky (2004, p. 729), who engage the topic of conflict resolution from a cognitive perspective. They focus on three relevant phenomena: ‘optimistic overconfidence, the certainty effect, and loss aversion’.

#### *3.4a Optimistic overconfidence.*

Overconfidence in agent judgements is ‘indicated by a cluster of robust findings’, essentially assignments of probability that are objectively unreasonable given the agent’s actual knowledge. According to Kahneman and Tversky (2004, p. 732), the inability to appropriately judge missing or indeterminate information ‘entails a bias that is likely to hinder successful negotiation’. The effect is ‘not restricted to laboratory studies’ and the study provides references of ‘real world data’ pointing to the same phenomenon. How are agents to reconcile whether their own confidence in a judgement is founded, or overstated, if those they intersubjectively compare with are in a similar psychological situation?

#### *3.4b Certainty effect.*

Research on decision making has ‘identified a major bias in the weights that are assigned to probabilistic advantages and to sure things’. A phenomenon most widely known as the Allais paradox indicates that people are not normatively rational in the way they view matters of chance, choices can be, often dramatically, at odds with the objective utility associated with agents’ subjective probability assertions (Kahneman and Tversky, 2004, pp. 734-735). For deliberation, and the discussion of means-ends comparisons, the implications are obvious. ‘Most decisions [made]...under uncertainty...involve vague contingencies and ambiguous probabilities. The evidence suggests that the certainty effect is further enhanced by vagueness and ambiguity’ (Kahneman and Tversky, 2004, p. 735). Habermasian agents, it would seem, cannot avoid operating in environments of indeterminateness, particularly given the requirement for intersubjectivity.

#### *3.4c Loss aversion.*

This last conflict related phenomenon speaks to the intuitive concerns of earlier critics who wondered whether Habermas expected too much of his agents; whether he forgot their humanity. Loss aversion is an empirically robust finding that a cognitive bias exists ‘toward the retention of the status quo.’ Individuals act in a manner that is not logically

rational in wanting to hold what they have, instead of trading it for something that is empirically and transparently of higher utility (Kahneman and Tversky, 2004, p. 742).

As Kahneman and Tversky study conclude, the three above phenomena ‘represent systematic departures from the standard rational theory ... these biases in the assessment of evidence and the evaluation of consequences can hinder the successful resolution of conflict...’ And the phenomenon might not be an exhaustive survey of conflict related psychological considerations, ‘they represent serious obstacles that often stand in the way of successful negotiation’. For Habermas, the conclusion runs even deeper when the authors state finally that the ‘...literature on judgement and choice ... indicates that biases and cognitive illusions are not readily eliminated by knowledge or warning’ (Kahneman and Tversky, 2004, pp. 743-44). Habermas’s assertion that the elimination of agent-relative concerns, through discussion, informational improvements or better argumentation, runs counter to such empirical evidence.

### *3.5 The preference problematic for Habermasian agents*

Kahneman (200, p. 758) asks: if the history of an individual through time ‘can be described as a succession of separate selves...[which] of these selves should be granted authority over outcomes that will be experienced in the future?’ Kahneman (2000, p. 758) brings into focus an issue that strikes at the heart of economic theory, utility theory, decision theory, and without doubt, deliberative theory and communicative rationality:

The expression of preference by means of choice and decision making is the essence of intelligent, purposeful behaviour. Although decision making has been studied for centuries by philosophers, mathematicians, economists, and statisticians, it has a relatively short history within psychology.

Agents must have preferences, as without them, they have no motivation to argue, to take positions, to test the validity of claims. Agents must presume to want one thing more than another, indeed, they must value one thing more than another. It is clearly not possible here to do justice to the voluminous literature that questions preference theory in either conceptual or empirical terms; it is though worth offering some brief remarks drawn from empirical psychologists. As mentioned earlier, there is evidence that differing elicitation procedures draw different choices from agents. Of this phenomenon, Paul Slovic (2000, p. 490) asks: ‘If different elicitation procedures produce different orderings of options, how can preferences be defined and in what sense do they exist?’ Much of preference theory tends to turn on notions of utility maximizations, and WARP (weak axiom of revealed preference), but as Slovic (2000, p. 489) contends ‘It is now generally recognized among psychologists that utility maximization provides only limited insight into the processes by which decisions are made ... Preferences appear to be remarkably labile, sensitive to the way a choice problem is described or “framed” and to the mode of response used to express the preference.’ Recent psychological research is ‘particularly troubling’ to classical and neo-classical economists because their theories are built on rational agents who hold preference sets that are both complete and transitive, i.e. ‘they choose what they most prefer’ (Slovic, 2000, p. 491). Yet as the observations in the

preceding sections demonstrate, choices may be capricious, and the only reliable phenomenon is that people will seek compelling rationalizations to explain why they chose the way they chose. Habermas, as a sociologist and philosopher, is remiss in not mitigating for such phenomenon. ‘The fact that preferences are highly labile, which psychologists have worked so hard to demonstrate, has been known to practical philosophers for ages’ (Slovic, 2000, p. 502).

### *3.6 Rationality – careful conclusions*

If we feel comfortable accepting that agents are ‘often controlled by emotions and desires that do not fit the model of calculating rationality’, that individual agency is ‘bounded by limitations on memory and computational capabilities’ and that the ‘experimental analysis of inference and choice has revealed that ... human judgement and decision making is often inconsistent with the maxims of rationality’ then we might see problems with deliberative outcomes, and universal assumptions of communicative rationality (Quattrone and Tversky, 2000, p. 452). Rationality sceptics might assume this means that human agency is completely capricious and irrational, that such questioning takes us one step too far. Indeed, certain normative principles do not seem to hold in certain circumstances, such as invariance and coherence; the safe conclusion is merely ‘that judgement and choice – like perception and memory – are prone to distortion and error.’ The significance here is that errors are common and systematic, not idiosyncratic and random, and thus ‘there is little hope for a theory of choice that is both normatively acceptable and descriptively adequate’ (Quattrone and Tversky, 2000, p. 472). Habermas’s expectations of his agents are both conceptually and psychologically heroic. While he may not have attended sufficiently to the available empirical research at the time of outlining his preconditions for communicative rationality, such a defence remains untenable for contemporary theorists. After all, the purpose of social science is to provide better theories, with more predictive explanations of social behaviour. Although, admittedly, this survey of empirical findings is not exhaustive, it is nevertheless instructive. Quite simply, and worryingly for Habermas:

voluntary action can proceed without accompanying conscious intending. As recent research demonstrates, thoughts, feelings, and behaviour that carry features of goal-directedness can emerge directly without a person’s conscious intent (Maasen, Prinz, and Roth, 2003, p. 103).

## **4. Conclusion**

We conclude this paper by pre-empting a major objection. The objection will come from social psychologists active and optimistic in the domain of deliberation research: if one looks at deliberative polling research where individuals are tested on their opinions before and after a deliberative and informational programme, one can draw positive conclusions. Researchers are comfortable determining that deliberation results provide on aggregate, ‘a better informed and more thoughtful public opinion’ (Luskin, Fishkin and Jowell, 2002, p. 484). The data is encouraging for deliberation as a forum. The evidence,

robust and compelling, indicates that people will change their preferences, will become more knowledgeable, and are quite likely to surrender their prior preference orderings under the weight of the better argument. Yet the research is ‘not what might be expected of an ideally informed and thoughtful citizenry’ (Luskin, Fishkin and Jowell, 2002, p. 484). The ideal citizenry is not the place to critique these deliberative results. Whether or not there is a superior political outcome post-deliberation is immaterial to this thesis. Deliberative polling requires sessions with experts, sessions with politicians, and most significantly, it requires administrators who go into the programme with the agent-centric goal of achieving deliberation. Habermas allows none of these, and thus positive deliberative results from these programmes are not supportive of Habermas’s requirements. The heroic preconditions of the Habermasian agent means deliberation must occur in a symmetrical, non-coercive, naturally occurring forum where there is no ‘expert’ status assigned to particular members of the deliberative forum and only the force of better arguments. As Dryzek (2000, p. 162) notes: the ‘authenticity of deliberation requires that communication must induce reflection upon preferences in a non-coercive fashion.’

The danger involved in taking Habermas’s earlier approach to communicative action and applying it in literal fashion to deliberative politics is that we end up with a political theory that has little to say about political structure – except to condemn it as an agent of distortion. For under communicative rationality – especially in its counterfactual extreme of the ‘ideal speech situation’ – the only force that applies is that of the better argument. Decision is ideally secured by consensus; implementation of the decision is secured only by the commitment of the individuals involved to the content of the consensus; and subsequent compliance relies on free consent. Such a sequence is not easily related to real-world political institutions and processes, especially those in complex and plural societies (Dryzek, 2000, p. 24).

The Habermasian agent is clearly not meant to operate in the environment of deliberative polls. Whether or not theorists and psychologists want to support deliberative forums, the agent within these forums is not the Habermasian agent that adheres to the conditions outlined above.

Ideally new research programmes could point more precisely to the deficiencies of Habermas’s rationality requirements. Where empirical data currently needs to be abstracted, novel research programmes like ethnographies in community settings could directly test the effectiveness of Habermasian-required communicative rationality assumptions. Our prediction, as deflationary as this seems, is that for deliberative success to ensue, groups need experts to mitigate and reconcile the psychological and informational phenomenon outlined above. Yet, as mentioned in the introduction, this incurs a Habermasian paradox. To attain the institutional requirements of successful deliberation one must break from Habermas’s procedural preconditions concerning power and communicative symmetry. To ensure deliberative success, non-neutral agendas are required, and expert communication is demanded. The conclusion for this paper is an opening for further study. Deliberative democracy has freshly problematized the nature of communicative rationality; successful deliberation requires heroic agents. Habermas has always demanded exceptionally stringent conditions for agents in communicative interactions, and for communicative rationality these conditions are not idealisations. Habermas sees the heroic elements of communicative rationality as given in



his agents. Contemporary psychological research however, makes the story a difficult one to accept. Although humans are not manifestly irrational, a full range of psychological phenomenon supports the deliberative theorists problems with Habermasian agents. Habermas should expect neither normative nor descriptive validity for the agents of his communicative rationality.

## Notes

\* Byron Rienstra, [GET INFO](#)

\*\* Derek Hook, Institute of Social Psychology, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE. Email: D.W.Hook@lse.ac.uk

## References

- Bohman, James (1997). 'Deliberative democracy and effective social freedom: Capabilities, resources, and opportunities', in: James Bohman and William Rehg (Eds.), *Deliberative Democracy* (Cambridge, M.A.: MIT Press).
- Dryzek, John S. (2000). *Deliberative Democracy and Beyond: Liberals, Critics, Contestations* (Oxford: Oxford University Press).
- Elster, Jon (1997). 'The market and the forum', in: James Bohman and William Rehg (Eds.), *Deliberative Democracy* (Cambridge, M.A.: MIT Press).
- Elster, Jon (1998), 'Introduction', in: Jon Elster (Ed.), *Deliberative Democracy* (Cambridge: Cambridge University Press).
- Gutman, Amy and Dennis Thompson (2004). *Why Deliberative Democracy?* (Princeton: Princeton University Press).
- Habermas, Jürgen (1982). 'A reply to my critics', in: John B. Thompson and David Held (Eds.), *Habermas: Critical Debates* (London: Macmillan).
- Habermas, Jürgen (1984). *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society* (Boston: Beacon Press).
- Habermas, Jürgen (1991). 'A reply', in: Axel Honneth and Hans Joas (Eds.), *Communicative Action* (Cambridge, Polity Press).
- Habermas, Jürgen (1993). 'Further reflections on the public sphere', in: Craig Calhoun (Ed.), *Habermas and the Public Sphere* (Cambridge, M.A.: MIT Press).
- Habermas, Jürgen (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (Cambridge: Polity Press).
- Heller, Agnes (1982). 'Habermas and marxism', in: John B. Thompson and David Held (Eds.), *Habermas: Critical Debates* (London: Macmillan).
- Johnson, James (1998). 'Arguing for deliberation: Some skeptical considerations', in: Jon Elster (Ed.), *Deliberative Democracy* (Cambridge: Cambridge University Press).

- Kahneman, Daniel (2000). 'New challenges to the rationality assumption', in: Daniel Kahneman and Amos Tversky (Eds.), *Choices, Values, and Frames* (Cambridge: Cambridge University Press).
- Kahneman, Daniel and Amos Tversky (2004). 'Conflict resolution: A cognitive perspective', in: Eldar Shafir (Ed.), *Preference, Belief and Similarity* (Cambridge, M.A.: MIT Press).
- Lukes, Steven (1982). 'Of gods and demons: Habermas and practical reason', in: John B. Thompson and David Held (Eds.), *Habermas: Critical Debates* (London: Macmillan).
- Luskin, Robert C.; James S. Fishkin and Roger Jowell (2002). 'Considered opinions: Deliberative polling in Britain', *British Journal of Political Science*, 32: 455-487.
- Maasen, Sabine; Wolfgang Prinz, and Gerhard Roth (2003). 'Voluntary action from the perspective of social-personality psychology', in: Sabine Maasen, Wolfgang Prinz, and Gerhard Roth (Eds.), *Voluntary Action: Brains, Minds, and Sociality*. (Oxford: Oxford University Press).
- March, James G. (1986). 'Bounded rationality, ambiguity, and the engineering of choice', in: Jon Elster (Ed.), *Rational Choice* (New York: New York University Press).
- McCarthy, Thomas (1982). 'Rationality and relativism: Habermas's overcoming of hermeneutics' in: John B Thompson and David Held (Eds) *Habermas: Critical Debates* (London: Macmillan).
- McCarthy, Thomas (1984). 'Translator's introduction', in: Jürgen Habermas, *The Theory of Communicative Action, Volume I: Reason and the Rationalization of Society* (Boston: Beacon Press).
- Pronin, Emily; Carolyn Puccio and Lee Ross (2002). 'Understanding misunderstanding: Social psychological perspectives', in: Thomas Gilovich, Dale Griffin and Daniel Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgement* (Cambridge: Cambridge University Press).
- Quattrone, George A. and Amos Tversky (2000). 'Contrasting rational and psychological analyses of political choice', in: Daniel Kahneman and Amos Tversky (Eds.), *Choices, Values, and Frames* (Cambridge: Cambridge University Press).
- Shafir, Eldar; Itamar Simonson and Amos Tversky (2000). 'Reason-based choice', in: Daniel Kahneman and Amos Tversky (Eds.), *Choices, Values, and Frames* (Cambridge: Cambridge University Press).
- Slovic, Paul (2000). 'The construction of preference' in: Daniel Kahneman and Amos Tversky (Eds.), *Choices, Values, and Frames* (Cambridge: Cambridge University Press).
- Sunstein, Cass R. (2003). 'The law of group polarization', in: James S. Fishkin and Peter Laslett (Eds.), *Debating Deliberative Democracy* (Oxford: Blackwell).
- Tversky, Amos and Daniel Kahneman (2000). 'Rational choice and the framing of decisions', in: Daniel Kahneman and Amos Tversky (Eds.) *Choices, Values, and Frames* (Cambridge: Cambridge University Press).
- Zajonc, Robert B. (1999). 'One hundred years of rationality assumptions in social psychology', in: Aroldo Rodrigues and Robert V. Levine (Eds.), *Reflections on 100 Years of Experimental Social Psychology* (New York: Basic Books).

## Endnotes

---

<sup>i</sup> Condorcet's Jury Theorem should claim a significant discussion here, but space does not allow.

<sup>ii</sup> Note, 'decentered' here is used in Piaget's sense, and Habermas explains the usage more fully than we are able to here.