Social Self-regulation in Computer Mediated Communities: The case of Wikipedia

Christopher Goldspink
Centre for Research in Social Simulation
Department of Sociology
University of Surrey. Guildford, GU1 7XH, UK
c.goldspink@surrey.ac.uk

Abstract

This paper documents the findings of research into the governance mechanisms within the distributed on-line community known as Wikipedia. It focuses in particular on the role of normative mechanisms in achieving social self-regulation. A brief history of the Wikipedia is provided. This concentrates on the debate about governance and also considers characteristics of the wiki technology which can be expected to influence governance processes. The empirical findings are then presented. These focus on how Wikipedians use linguistic cues to influence one another on a sample of discussion pages drawn from both controversial and featured articles. Through this analysis a tentative account is provided of the agent-level cognitive mechanisms which appear necessary to explain the apparent behavioural coordination. The findings are to be used as a foundation for the simulation of 'normative' behaviour. The account identifies some of the challenges that need to be addressed in such an attempt including a mismatch between the case findings and assumptions used in past attempts to simulate normative behaviour.

Introduction

The research documented in this paper is part of the EU funded project titled 'Emergence in the Loop: Simulating the two way dynamics of norm innovation' (EMIL) which aims to advance our understanding of emergent social self-organisation. The project involves conducting several empirical case studies the first of which is the Wikipedia.

When people encounter Wikipedia for the first time and learn how it works, they commonly express surprise. The expectation appears to be that an open collaborative process of such magnitude should not work. Yet the Wikipedia has been shown to produce credible encyclopaedic articles (Giles, 2005) without the hierarchical and credentialist controls typically employed for this type of production.

The research presented here is framed within the debate about governance mechanisms associated with Open Source production systems. This is not the only perspective which could be adopted but it does serve to provide some initial orientation. Consistent with the wider project focus, the relationship between these theories and the theory of social norms is examined.

In the empirical research we examine the extent to which communicative acts are employed by editors to influence the behaviour of others. Particular attention is given to the illocutionary force of utterances (Searle, 1969) and the effect of deontic

commands linked to general social norms and Wikipedia specific rules. In the conclusion some observations are made about the agent-level cognitive mechanisms which appear necessary to explain the observed social order as well as the apparent influence of social artefacts, goals and the wiki technology.

The following questions are canvassed through this research.

- What processes appear to operate in computer mediated organizations which enable them to be, in effect, self-regulating?

- How consistent are the findings with established theories for understanding norms and governance, particularly in on-line environments?

- What alternative hypotheses are there which appear to explain the phenomena and which can provide the foundation for future research?

Governance Theory

According to the relevant Wikipedia article, the word 'governance' derives from the Latin that suggests the notion of "steering". The concept of governance is used in a number of disciplines and a wide range of contexts and the range and type of steering mechanisms differ depending on whether the focus is with states or institutions. While both have been applied to Open Source, it is most common (and arguably most appropriate) to use institutional concepts of governance. Institutional steering mechanisms may be: formal (designed rules and laws) or informal (emergent as with social norms); extrinsic (involving contracts and/or material incentives) or intrinsic (involving values and principles); and the mechanisms by which governance operates may be top down (imposed by authority) or bottom up (invented by the participants as a basis for regulating each other). Theories vary with respect to the mechanisms advanced and the emphasis placed on different mechanisms. Theory is also advanced for different purposes: to explain or to prescribe. In broad terms the debate is often dichotomised with economics derived theories (Agency and Transaction Cost) on one side and sociological theories (stewardship) on the other (see J. H. Davis, D. Schoorman, & L. Donaldson, 1997; Donaldson & Davis, 1991). Depending on the position of the advocate these may be presented as antithetical or as viable alternatives for different contexts.

Agency theory derives from neo-classical economics and shares the foundational assumption of agent utility maximization. Advocates argue that many productive transactions involve *principals* who delegate tasks to *agents* to perform on their behalf (Donaldson & Davis, 1991). This gives rise to what is known as the 'principal's dilemma'. Simply stated this dilemma asks *'how can the principal ensure that the agent will act in its interest rather than on the basis of self-interest?'* Note that this dilemma arises from the assumed self-interested nature of agents –it is a dilemma intrinsic to the assumptions upon which the theory is based even though this is argued to have empirical support. Two general solutions are offered: the use of formal contracts and sanctions and the use of material incentives.

Critics argue that not all human decisions are made on the basis of self-interest. Sociological and psychological models of governance posit various alternatives: some remain committed to assumptions of rational action and goal seeking, while others address issues of power or various forms of intrinsic motivation, including a desire to conform to social norms. These latter positions generally form the basis of theories of *stewardship* (J. H. Davis, D. F. Schoorman, & L. Donaldson, 1997).

While these two broad sets of ideas form the backdrop to most debates about governance in traditional institutions increased recourse has also been made to Transaction Cost Economics (TCE) (Coase, 1993, 1995; Williamson, 1996). TCE is concerned with the relative merit of alternative governance arrangements for differing production environments. Oliver Williamson (1985), a key contributor, states *'The choice of governance mode should be aligned with the characteristics of the transaction…'*. Principals are presented with a continuum of possible ways of trying to achieve effective regulation from open markets to hierarchy. Both of these are seen as imposing costs (agency costs for hierarchy and transaction costs for markets). The aim is to combine them to achieve an optimum balance between these costs. This 'balancing' implies a top down rational decision making role for institutional managers.

More recently two additional categories of governance have been added to the TCE family – 'networks' and 'bazaars'. Both have arisen to explain the emergence of production and exchange arrangements which do not seem to fit on the market-hierarchy continuum. Both Network Governance (Candace Jones, William S Hesterly, & Stephen P Borgatti, 1997) and Bazaar Governance (Demil & Lecocq, 2003) are argued to be particularly relevant to understanding the flexible structures associated with Open Source production. Demil & Lecocq (2003: 8) cite Jones et al (1997: 916) and argue that network governance:

> *…involves a select, persistent, and structured set of autonomous forms [agents] engaged in creating products or services based on implicit and open ended contracts to adapt to environmental contingencies and to coordinate and safeguard exchanges. These contracts are socially – not legally – binding.*

The final sentence highlights the key difference between network and more conventional TCE mechanisms. To achieve cooperation the network form of governance relies on social control, such as *'occupational socialization, collective sanctions, and reputations'* rather than on formal authority.

Bazaar governance is also argued to rely heavily on the mechanism of reputation (Demil & Lecocq, 2003: 13). Reputation is assumed to provide the incentive to become involved and to comply with group expectations and norms. Unlike network systems, however, agents are free to enter or leave the exchange process – there are no obligations to become or to remain engaged. Raymond (1999) states that *'contrary to network governance, free-riders or opportunistic agents cannot be formally excluded from the open-source community'*.

To summarise:

- Free markets are characterised by: a lack of obligation to engage in a transaction; low interdependence between parties involved with the exchange; and transactions regulated only by price. Within a pure market the individual identities of the transacting parties are not important.

- In Hierarchies, there are formal contracted obligations on all parties, these are maintained by fiat but may also be supported by wider formal institutions e.g. Courts. Obligations are associated with formal position making the official (role) identity of the parties the key determinant of the relationship.

- Within network structures, exchanges are regulated using relational contracts – there is a formal obligation to remain engaged even though specific actions

and operational responsibilities may not be included in a contract. There is also some reliance on social norms– the socialised position of actors becomes important. Exchange commitments may be relatively short lived and persist only so long as they offer mutual benefit.

- With bazaar governance there is no obligation on any party to perform particular duties or even to remain engaged: there are low entry and exit costs. There are few formal mechanisms for policing or sanction but sufficient regulation is achieved by means of shared task, reciprocity norms and/or informal group sanctioning with participants influenced by their desire to build reputation.

Understanding the role of Norms

As can be seen, 'norms' are argued to play a role in a number of theories of governance, with their being particularly significant in Stewardship, Network and the Bazaar theories. Sociologists have long argued that norms are fundamental mechanisms for social regulation. What though is a 'norm'? How do norms emerge and how are they influenced and by what?

Gibbs (1981) argues that *'Sociologists use few technical terms more than norms and the notion of norms looms large in their attempt to answer a perennial question: How is social order possible?'*. Not surprisingly then the concept has been incorporated into a wide range of alternative and often competing bodies of theory.

The normative literature can be divided into two fundamentally distinct groups. In the social philosophical tradition (Lewis, 1969) norms are seen as a particular class of emergent social behaviour which spontaneously arise in a population. From this perspective, a 'norm' is a pattern identified by an observer ex-post. The defining characteristic of the pattern is the apparently prescriptive/proscriptive character: people behave 'as if' they were following a rule. By contrast, the view offered by the philosophy of law sees norms as a *source* of social order. This standpoint assumes the prior existence of (powerful) social institutions and posits them as the source of rules, which, when followed, lead to social patterns. These positions appear antithetical although following the work of Berger and Luckman (1972) each may be seen as a part of a dialectic whereby emergent social patterns become reintegrated and formalised in institutions.

Therborn argues (2002: 868) that people follow norms for different reasons. The extremes run from habit or routine to rational knowledge of consequences for self or the world. Between these lie:

- Identification with the norm or values – linking sense of self (identity) to the norm source (person, organization or doctrine) often leading to in-group-out-group.
- Deep internalization – self-respect – done independently to what others are doing.

Bicchieri (2006: 59) provides a rare hint at the cognitive process involved stating:

> To 'activate' a norm means that the subjects involved recognise that the norm applies: They infer from some situational cues what the appropriate behaviour is, what they should expect others to do and what they are expected to do themselves, and act upon those cues.

This suggests a complex process of self-classification (how am 'I' situated with respect to this group and what is the nature of the situation in which 'I' find myself, does a norm pertain to 'me' in this situation and under what conditions and to what extent am I obliged to comply?).

To begin to identify which (if any) of these loosely defined mechanisms might be supported by evidence and to aid in the development of a theory of norms helpful for understanding the more general mechanisms at play in social self-regulation we selected the Wikipedia as a preliminary case study. Wikipedia belongs to the Open Source movement as it has adopted the Open Source License. It was originally designed to operate under the umbrella of a conventional hierarchical form of governance and its unanticipated success as a radical governance experiment makes it a particularly interesting case study. It was anticipated that findings in relation to the Wikipedia may have some wider relevance to understanding the open source phenomena but also serve to cast light on mechanisms which underpin human institutions– particularly those that are more normative in nature. In order to be able to judge the degree of generalisation that may be possible it is first important to identify the distinctive features of the Wikipedia.

The Wikipedia

Wikipedia grew out of an earlier Web encyclopaedia project called Nupedia founded by Jimmy Wales with Larry Sanger appointed as its first editor-in-chief. From its inception Nupedia was linked to a free information concept and thus the wider open source movement. Nupedia used traditional hierarchical methods from compiling content with contributors expected to be experts. The resulting complex and time consuming process and an associated lack of openness have been argued to explain the failure of the Nupedia. Sanger (2006; 2007), however, questions this view, arguing that the expert model was sound but needed to be simplified.

Sanger was introduced to the WikiWiki software platform in 2001 and saw in it a way to address the limitations hampering Nupedia. The inherent openness of the Wikiwiki environment was, however, seen as a problem so Wikipedia began as an experimental side project. Sanger notes that a majority of the Nupedia Advisory Board did not support the Wikipedia, being of the view '*…that a wiki* could *not resemble an encyclopaedia at all, that it would be too informal and unstructured'* (Sanger, 2007). However the intrinsic openness of Wikipedia attracted increasing numbers of contributors and quickly developed a life of its own. Almeida et al (2007) note that growth in articles, editors and users have all shown an exponential trajectory. From Sanger's earlier comments it is clear that he had been surprised at the rate of development and of the quality achieved by the relatively un-coordinated action of many editors.

The Debate over Governance in Wikipedia

The use and enforcement of principles and rules has been an ongoing issue within the Wikipedia community with a division emerging between the founders and within the wider community about whether rules were necessary and if they were, how extensive they should be and how they should be policed. The power to police rules or impose sanctions has always been limited by the openness of the technology platform. Initially Sanger and Wales, were the only administrators with the power to exclude participants from the site. In 2004 this authority was passed to an Arbitration

Committee which could delegate administrator status more widely. The Arbitration Committee is a mechanism of last resort in the dispute resolution process, only dealing with the most serious disputes. Recommendations for appointment to this committee are made by open elections with appointment the prerogative of Wales.

In the early stages Sanger argues the need was for participants more than rules and so the only rule was 'there is no rule'. The reason for this, he explains, was that they needed to gain experience of how wikis worked before over prescribing the mechanisms. However, *'As the project grew and the requirements of its success became increasingly obvious, I became ambivalent about this particular "rule" and then rejected it altogether'* (Sanger, 2007). However, in the minds of some members of the community, it had become 'the essence' of Wikipedia.'

In the beginning, complete openness was seen as valuable to encourage all comers and to avoid them feeling intimidated. Radical collaboration – allowing everybody to edit everyone's (unsigned) articles – also avoided ownership and attendant defensiveness. Importantly it also removed bottle necks associated with 'expert' editing. That said the handpicking of a few core people is regarded by Sanger as having had an important and positive impact on the early development of Wikipedia. Sanger argues for example *'I think it was essential that we began the project with a core group of intelligent good writers who understood what an encyclopaedia should look like, and who were basically decent human beings'* (2005). In addition to 'seeding' the culture with a positive disposition, this statement highlights the potential importance of establishing a style consistent with the Encyclopaedia genre – a stylistic model which might shape the subsequent contributions of others.

Sanger argues that in the early stages 'force of personality' and 'shaming' were the only means used to control contributors and that no formal exclusion occurred for six months, despite there being difficult characters from the beginning. The aim was to live with this 'good natured anarchy' until the community itself could identify and posit a suitable rule-set. Within Wikipedia rules evolved and as new ones were needed they were added to the 'What Wikipedia is not' page'. Wales then added the 'Neutral Point of View' (NPOV) page which emphasised the need for contributions to be free of bias. The combination of clear purpose and the principle of neutrality provided a reference point against which all contributions could be easily judged. Sanger regards the many rules, principles and guidelines which have evolved since as secondary and not essential for success.

How do newcomers learn these (ever increasing) rules and do they actually influence behaviour? Bryant et al (2005) suggest that there is evidence of 'legitimate peripheral practice', a process whereby newcomers learn the relevant rules, norms and skills by serving a kind of apprenticeship. These authors argue that this is evident in new editors of Wikipedia initially undertaking minor editing tasks before moving to more significant contributions, and possibly, eventually, taking administrative roles. These authors tend to project a rather idealistic view of involvement, however, overlooking a key attribute of the wiki environment –newcomers have the same rights as long standing participants and experts and this mechanism for socialising newcomers can be effectively bypassed.

In some Open Source environments (such as Open Source Software) it is possible to gain reputation which may be usable in the wider world. The commitment to the community is often explained (for an excellent overview see Rossi, April, 2004) by arguing that a desire for reputation increases compliance. However, in the Wikipedia

environment there is no list of contributors to which an editor can point as evidence of their contribution (although they can self-identify their contributions on their user page). Contributions are, in essence, non attributable. In the case of Wikipedia identification with product, community and values appears a more likely reason for remaining involved than does reputation.

In a study specifically designed to study the conflict and coordination costs of Wikipedia, Kittur, Suh, Pendleton, & Chi (2007: 453) note that there has a been a significant increase in regulatory costs over time. *'...direct work on articles is decreasing, while indirect work such as discussions, procedure, user coordination, and maintenance activity (such as reverts and anti-vandalism) is increasing'*. The proportion of indirect edits (i.e. those on discussion or support pages) has increased from 2% to 12%. Kittur et al cite an interview respondent as stating *'the degree of success that one meets in dealing with conflicts (especially conflicts with experienced editors) often depends on the efficiency with which one can quote policy and precedent.'* (Kittur et al., 2007: 454). This suggests that force of argument supported by the existence of the formal rules and etiquette are important to the governance process. This is however based on ex post attributions.

Wiki technology- the artefact

Wiki technology has a **very flat learning curve**: contributing is extremely simple. There are few technical impediments confronting novice users. Wiki platforms are **intrinsically open** supporting decentralised action unless modified to control or restrict access. **Division of labour emerges** as editors choose which pages interest them and which they want to focus on contributing to or maintaining.

Wikipedia has added a number of facilities which support the ready detection and correction of vandalism. **Watch lists** support users in taking responsibility for the oversight and monitoring of particular topics. Changes made to a page are logged using a **history list** which supports comparison between versions as well as identifying the time and date of any change and the ID of who made that change. The **reversion** facility supports the rapid reinstatement of the page content. Lih (2004: 4) attributes significance to this feature noting that *'This crucial asymmetry tips the balance in favour of productive and cooperative members of the wiki community, allowing quality content to emerge'.* and Stvilia et all (2004: 13) note that *'By allowing the disputing sides to obliterate each others contributions easily, a wiki makes the sides interdependent in achieving their goals and perhaps surprisingly may encourage more consensus building rather than confrontation'.*

Stvilia, Twidale, Gasser, & Smith (2005) among others identify **discussion pages** as an important *'...coordination artefact which helps to negotiate and align members perspectives on the content and quality of the article.'* Discussion pages provide an opportunity for managing minor disputes about content or editing behaviour and for movement towards the agreement.

Ciffiolilli (2007) has argued that a significant consequence of these technical features is the way in which they alter transaction costs (Coase, 1993; Williamson & Winter, 1993). Transaction costs result from information overheads associated with complex coordination. However, the technology does not cancel other costs of coordination and control. These are commonly referred to as agency costs and the highly open nature of the wiki may increase them. In hierarchies, this cost is evident in the cost of command and control (management overhead) whereas in the Open Source

environment they are borne by the participating community (and not necessarily equitably). The cost burden will be less where there is a high level of self-regulation and lower where a lack of goal alignment or low social commitment leads contributors to disregard others and act individualistically or opportunistically. The efficacy of cultural control will be influenced by factors such as the homogeneity of the user group and that group's propensity for self-organisation (endogenous norm formation), rates of turnover of the group, and the effect of external perturbation of the group or of the task on which they are working. This may also be subject to feedback effects: reduced norm compliance may lead to higher turnover and reduced commitment, further reducing norm compliance for example.

In conclusion then, Wikipedia is a volunteer open source project characterised by low ties between contributors, no formal obligations and very few means for the exercise of formal sanction. There is a low level of reciprocity with contributors under no obligation to maintain engagement. The wiki technology is open, inviting many to the task and imposing low costs to participation while reducing transaction costs. There is however high reliance on pro-social behaviour dominating if agency costs (borne by individuals) is not to lead to high turnover and possible governance failure. The anonymity of Wikipedia precludes any significant reputation effects outside of the small group of co-editors who maintain extended involvement with an article and to a very limited degree the wider Wikipedia community.

Wikipedians have produced a set of permissions, obligations, rules and norms which have been documented in guidelines and etiquettes as well as embedded in technical artefacts such as style bots. The need for and effect of these is however controversial. From a governance perspective there are relatively few means within Wikipedia by which formal control can be exercised using these rules and the community relies instead on the use of informal or 'soft' control. These mechanisms need to be effective in the face of perturbation from 'vandals' (task saboteurs), 'trolls' (social saboteurs), as well as turnover of contributors in the context of a task which can require the accommodation of emotionally charged and value based issues.

Analysis of Governance Micro-mechanisms.

In Wikipedia there are two classes of activity: editing; and conversation about editing. This paper is not concerned with the editing activity (although this is to be considered in future research) but with the self-organising and self-regulating phenomena which make it possible. Insight into this can be gained by examining the Discussion pages which accompany many of the articles rather than the articles themselves. The activity on the Discussion pages comprises a series of 'utterances' or speech acts between contributors about editing activity and the quality of product. On the face of it then, these pages should provide a fertile source of data to support analysis of how governance operates in the Wikipedia, in particular informal or 'soft' governance.

Within these pages we expected to see attempts by editors to influence the behaviour of one another through the only means available to them – communicative acts. We anticipated that these may exhibit some regularity which would allow us to examine both the range and type of events that led to the explicit invocation of rules and norms and which revealed emergent influence patterns which were themselves normative. We wanted also to examine what conventions prevailed and how these compared and interacted with the goal of the community and its policies. A convention is defined here as a behavioural regularity widely observed by members of the community.

Policies include explicit codes of conduct as well as guidelines (etiquettes) and principles.

Methodology

For the study we randomly selected a sample of Discussion pages associated with both Controversial and Featured articles. At the time of the study (May/June 2007) there were 583 articles identified by the Wikipedia community as controversial. The featured articles are more numerous. At the time of the study there were approximately 1900 of them. The analysis reported here is based on a sample of nineteen Controversial and eleven Featured articles. The most recent three pages of discussion were selected for analysis from each Discussion page associated with the article included in the sample.

These were subjected to detailed coding using the Open Source qualitative analysis software WeftQDA. Both qualitative and quantitative analysis was performed. The latter was undertaken by re-processing the coded utterances such that each utterance constituted a case and each applied code a variable associated with that case. This data set was then analysed using SPSS.

A number of coding schemes for natural speech were considered before choosing to use the Verbal Response Mode (VRM) taxonomy (Stiles, 1992). VRM has been developed over many years and used in a wide range of communication contexts. Stiles defines it as *'a conceptually based, general purpose system for coding speech acts. The taxonomic categories are mutually exclusive and they are exhaustive in the sense that every conceivable utterance can be classified.'* (Stiles, 1992: 15). The classification schema is attractive where there is a need (as here) to capture many of the subtleties of natural language use that derive from and rely on the intrinsic flexibility and ambiguity of natural language yet map them to a more formal system needed for computer simulation.

Additional codes were applied to identify: valence, subject of communication, explicit invocation or norms or rules and the associated deontic and trigger, whether the receiver/s accepted the illocutionary force of the utterance and the ID and registration status of the person making the utterance.

There were 3654 utterances coded in these thirty three documents.

Findings

Style of Communication

There was a statistically significant correlation between the article group (Controversial vs Featured) and broad style of communication. This was however very small at -0.078 (p=.01 2-tailed). This difference was most apparent when examined at the level of specific styles. Both groups had approximately similar proportions of neutrally phrased utterances (approximately 64%). Nearly one quarter (22.5%) of all utterances in Featured articles were positive compared to only eleven percent in controversial sites. By comparison nearly one quarter (23.9%) of all utterances in controversial sites were negative compared to fourteen percent for featured. The positive styles of 'affirming', 'encouraging' and 'acknowledging' were significantly overrepresented in the featured articles but underrepresented in the controversial articles. The reverse was the case for the negative styles of 'aggressive', 'contemptuous' and 'dismissive'.

There was a statistically significant correlation between the broad style of communication and the editor status. The correlation was again very low at -.054 (p=.01 two tailed).

Overall, the most common positive utterance was affirming (4.7%) closely followed by encouraging (4.7%) and acknowledging (4.3%). The most common negative utterance was dismissive (8.2%) followed by defensive (6.4%) and contemptuous (3.5%).

All the Wikipedia discussions sampled reflected a strongly neutral-objective *style* (although from the qualitative observations it was apparent that the content was sometimes far from objective or balanced). The statistically significant difference between Controversial and Featured sites was in the relative balance of positive and negative utterance and was not so great as to explain the different status awarded the associated articles.

Validation

Within speech act theory (Habermas, 1976; Searle, 1969), validation refers to whether an utterance made by one speaker is accepted, rejected, ignored or let go unquestioned by the intended recipient/s.

In the Wikipedia sample half of all utterances were accepted without question. A further eighteen percent were explicitly accepted by at least one editor; eleven percent were explicitly rejected and a substantial twenty two percent were ignored. Twenty five percent of positive style utterances were accepted by at least one editor compared to eighteen percent of neutral and only nine percent of negative. By comparison only two percent of positive utterances were rejected compared to nine percent of neutral and twenty six percent of negative. Positive utterances were more likely to be accepted without question (61%) compared to negative (21.7%) and neutral (54.4%). Negative comments were more likely to be ignored (44.1%) compared to neutral (18.2%) and positive (11.4%).

From this we can conclude that positive utterances are more likely to be validated than negative, but that overall, a significant number of utterances are ignored or rejected.

Normative and rule invocation

Overall 5.2% of all utterances involved norm or rule invocation. This meant that Wikipedia rules were invoked 122 times and general social norms a further 77 times in 3654 utterances. This overall number was contributed to disproportionately by three (outlier) articles in the sample. Rules were most commonly invoked in response to neutral style communication (63.9%) followed by twenty seven percent in response to a negative style. Only nine percent of positive style utterances were responded to with a rule invocation. By comparison, norms were most commonly invoked in response to negative style utterances (53.2%) followed by neutral (44.2%) and then positive (2.6%). The difference in likelihood of invocation by style was statistically significant (p=.001).

A Wikipedia rule invocation was most likely to be triggered by the *form* of an article (44.9%) an *edit action* (22%); an *article fact* or a *person's behaviour* (both 16%). A norm was most likely to be triggered by a *person's behaviour* (35.6%), an *edit action*

(23.3%), *article form* (21.9%), or *article fact* (19.2%). This pattern did not differ to a significant degree between the Featured and Controversial sites.

Nearly three quarters (73.6%) of rule invocations had the implicit deontic of 'it is obligatory' Norms also were most likely to carry this deontic (61.3%). The second most likely deontic was 'it is permissible that' (9.7%).

While there was no statistically significant difference in the degree to which either norms or rules were invoked between the Featured and Controversial articles, there was a qualitative difference in the role norm and rule invocation played. In Controversial discussions, social norms and rules were most likely to be invoked against the behaviour of an editor who was of a different view (group?) while in Featured sites, norms and rules were somewhat more often used by the editor as a reflection on their own contribution – i.e. involved a level of self-check. This might take the form of a statement such as 'I know this is not NPOV but…..'.

Registered vs non-registered users

There was no statistically significant difference in the likelihood for either registered or non-registered users to invoke norms or rules. There was a statistically significant difference between registered and non-registered editors (p=.000) when it came to validation. Registered editors were more likely than non-registered to be explicitly accepted (18.7% of utterances compared to 13.9%), less likely to be rejected (9.9% compared to 13.7%), considerably less likely to be ignored (18.3% compared to 34.7%) or unquestioned (53.1% compared to 37.6%). Qualitatively, however, it was much more common that un-registered users would make suggestions before undertaking edits, particularly in the Features articles, so their behaviour was less likely to attract action or comment.

Non-registered editors were more likely to make negative style utterances (24.3% compared to 18.5%) and less likely to make positive utterances (9.5% compared to 17.4%). This difference was significant (p=.000).

Influence through Illocutionary Force

The theory of speech acts distinguishes between the meaning of an utterance and its pragmatic intent. With the VRM coding frame used in this research each utterance is coded twice, once to capture the semantic form and again to capture the use of language to exert (illocutionary) force (Searle, 1969). A typical utterance may have a *form* which differs from the *intent*. The utterance 'could you close the door?', for example, has the form of a *question* but the intent of *advisement:* the speaker intends the listener to close the door. In VRM, the relationship of form to intent is expressed, using the statement "in service of" (Stiles, 1992). In this example the question 'could you close the door' is 'in service of' the advisement 'close the door'. In standard presentation this is recorded as (QA).

Edification in service of Edification (EE) is the most frequent form of utterance in the Wikipedia sample – 37% of all utterances were of this mode. The Edification mode is defined as deriving from the speaker's frame of reference, making no presumption about the listener and using a neutral (objective) frame of reference shared by both speaker and listener. This mode is informative, unassuming and acquiescent. As a strategy for influencing others it reflects attempts to convince by neutral objective argument.

The second most common mode is that of Disclosure in service of Disclosure (DD). Disclosure is defined as being from the speaker's experience, making no presumption, but being framed using the speaker's frame of reference. This is summarised as informative, unassuming but directive. Unlike EE mode, DD mode represents an attempt by the speaker to impose or have the listener accept the speaker's frame. Twelve percent of all utterances adopted this form.

The third most common mode is Disclosure in service of Edification (DE). The DE mode represents an utterance which is from the speaker's frame of reference but as if it is neutral or from a shared frame. Eight percent of all utterances used this mode. This is a somewhat neutral mode where the speaker offers clearly labelled personal knowledge as information.

The fourth most common mode is Advisement in service of Advisement (AA). AA mode represents speech from the speaker's experience, which makes presumptions about the listener and adopts the speaker's frame of reference. It can be summarised as informative, presumptuous and directive. It commonly takes the form of 'you should….' Approximately 7% of utterances were in this mode. A further 12% of utterances have the directive pragmatic intent of advisement masked by using a less presumptuous form – that of Edification or Disclosure.

Significantly, utterances associated with politeness (such as acknowledgements 5%) and with discourse which aims at mutual understanding, such as confirmation (1.5%) and reflection (1%), were very rare in the Wikipedia sample.

Discussion of Findings

What is significant about the utterance strategies is that they typically involve an exchange of assertions delivered with a neutral – i.e. non-emotive style. There are very few explicit praises, or put downs, and few niceties like explicit acknowledgements of one another. Seldom do contributors refer to one another by name – the exchanges are rather impersonal. This does not tally with what one would expect if the Wikipedia etiquette (http://en.wikipedia.org/wiki/Wikipedia:Etiquette) had been institutionalised. The Featured articles conform a little more closely with what one would expect than do the Controversial, but if we assume that the etiquette captures the community's ideal, the emerged patterns do not conform to that 'ideal' to the extent that might be expected in either case. Similarly we see low levels of questioning or of reflection (i.e. feeding back the words of the speaker to check understanding or to come to better understand the other's intentions). This is arguably inconsistent with the task needs – to reach consensus on controversial topics. The frequency with which utterances were ignored also suggested low engagement by participants in the discussion. All of this would seem to need some explanation.

The absence of any expression of acknowledgement of emotions and/or similarity of attitude (homophilly) among many contributors suggests that Wikipedia lacks many of the qualities of verbal exchange that would identify it as strong community. It is more consistent with being a place to share coordination of a task. This could suggest that the goal is the primary orientating point. However, the lack of quality of discourse needed to achieve consensus is more indicative of a brief encounter between different and established milieux which struggle to find common understanding rather than of a community committed to a common goal (Becker & Mark, 1997). This might suggest that the shared goal may be subordinate to more personal goals by a

considerable proportion of contributors. Or it may be that the technology and environment will support no more than this.

The Wikipedia environment supports saboteurs who can use the opportunity afforded by the open and anonymous platform to use identity deception i.e. to mimic the language and style of an 'expert' or to present as a genuine editor while trying to pursue a personal or political agenda hostile to the aims or interests of the Wikipedia. We found no direct evidence of this behaviour in the pages we sampled even though the discussions about controversial articles provide particularly fertile ground for such sabotage. Nevertheless the threat of it could have an overall influence on the type of communication conventions which arise. Editors may, for example, display reserve and suspicion, withholding trust and taking conventional signals of authority and identity (Donath, 1998) as unreliable. The first principle in the Wikipedia etiquette is 'assume good faith'. To do so would, however, leave the process more vulnerable to 'troll' activity.

Utterance strategies between registered and unregistered editors did not vary greatly, although unregistered editors were more likely to use disclosure intent and more likely to ask questions (possibly associated with the increased likelihood that they are relatively new to Wikipedia). They are also more likely to be negative – reflecting their potentially lower commitment to the article or the community.

Qualitatively there was considerable evidence of mind reading (theory of mind) – i.e. editors appeared to form judgements about the intent of others on relatively little information. There was, however, little evidence of the use of utterance strategies to better understand or check these theories of mind. Some editors, particularly in the Controversial discussions appeared quick to judge and then follow response patterns consistent with those judgements (e.g. ignoring or accepting utterances of others). There were also few instances of renegotiated patterns of communication style. Positions and styles stayed relatively constant over the period of the interaction. Only occasionally would an editor modify his/her style significantly if challenged. Of the rule invocations 26% were accepted, a similar proportion were rejected or ignored and the remainder went unquestioned (but generally had no affect on behaviour). This is consistent with norms being triggered by a limited range of cues which allow individuals to locate themselves and select identities appropriate to a context and which then remain essentially stable. The invocation of rules and norms appears to have little to no immediate effect on behaviour although it is not clear if it has an effect in subsequent behaviour as this cannot be ascertained from the available data.

Conclusions and future work

In this study we set out to identify mechanisms which underpin the emergence of systemic self-organisation in a volunteer on-line global institution. The aim was to specify the mechanisms involved in order to support the design of a simulation architecture suitable for the wider study of normative mechanisms. The findings have challenged some of our assumptions and expectations, in particular:

- The more detailed and specific behavioural etiquette seems to have little influence on the overall character and style of interaction.

- The overall quality of interaction of editors falls short of the range and quality of communicative style characteristic of a community and that would be consistent with what one would expect, given the nature of the task.

- Most regulation is achieved without the need for frequent explicit invocation of rules or norms. Rather, behaviour seems to accord to a convention which editors quickly recognise and conform to (or bring to the Wikipedia) and which minimally accommodates what needs to be done to satisfy the task in a context of potentially heterogeneous personal goals.

- There was a lack of evidence of active negotiation of expectations and standards and convergence of behaviour towards a norm. Within the discussion pages there appeared to be little obvious norm innovation, evolution, adaptation or extension. This suggests that on first encounter with Wikipedia, editors read a set of cues as to what constitutes appropriate or acceptable behaviour and then accommodate it. Alternatively the order observed may be largely attributable to the prior socialisation of participants with local norms and rules playing a very minor part in supporting task regulation.

- While there is a difference between controversial and featured sites this is minimal and the quality of the interaction cannot explain the difference in status. Similarly there appeared to be little in the subject matter of the two groups of articles which would explain the difference – both contained subject matter which was contestable and subject to significantly diverse opinion.

Wikipedia is not a market as there is no tradable product or price, either in a conventional sense or in the form of tradable reputation. Nor is Wikipedia a command hierarchy: the openness of the wiki platform and the low cost of joining and leaving precludes formal control as a primary means for governance. Neither is Wikipedia well described by the network theory of governance as there is no obligation to maintain involvement. While it might be expected that the Bazaar Governance would apply, the absence of a reputation mechanism suggests that it may be better considered through the more general lens of stewardship theory. Even here, there is no role for moral leadership but rather a diffused willingness to comply with certain minimum standards on the part of a sufficient majority.

There is no clear basis to argue that the apparent order is a direct result of the use of deontic commands associated with social norms and environment specific rules. Despite the fact that the community has been a prolific rule generator, they appear to play a minor role. Contributors demonstrate a style which is broadly inconsistent with these rules and not a good fit with the task.

Overall though there is order and it appears to be emergent. The mechanisms which underpin this emergence have not been revealed by the analysis undertaken to date although some hypotheses can be tentatively suggested. The neutral-objective style may be a consequence of the anonymity and open nature of the environment – leading to a suspension of trust. It may propagate as new comers copy the pattern through a process of behavioural cueing. It is possible also that the order is due to pro-social behaviour internalized and brought to the task. The volunteer nature of Wikipedia, and the level of commitment required, is likely to mean that long term editors reflect a pro-social disposition (Penner, Dovidio, Piliavin, & Schroeder, 2005). In this context a little norm/rule invocation may go a long way if not by influencing immediate behaviour then by encouraging future compliance and/or by giving incentive for non-compliers to leave. The relatively small difference in overall style apparent in relation to the diverse range of articles may have little to do with the specific communicative behaviours adopted in communication about that article but rather due to the chance

association of individuals at a given point of time and how this subtle process of encouragement and dissuasion plays out over time. Such a view is quite different from that modelled in past attempts to simulate social norms.

A review of past approaches to the simulation of norms undertaken by EMIL partners at the University of Bayreuth concluded that the past research drew on the traditions of game theory and artificial intelligence. The latter were exclusively in the first generation AI tradition. Significantly, data drawn from real social situations was seldom used and there was a strong tendency to build on prior work with little questioning of assumptions about the nature of normative behaviour. Seldom was any mainstream theory of social behaviour employed as a part of the research program. The EMIL project is notable, therefore, for its insistence on the need to adopt an empirical orientation: for models to be designed in the light of and tested against observations drawn from real world cases of normative behaviour as well as in its avoidance of pre-commitment to particular simulation models or traditions.

While the findings of the research to date are far from conclusive they do challenge many of the assumptions incorporated into past simulations and suggest a range of alternative hypotheses. Some of these will be able to be critically examined by further analysis of the current data and/or by data currently being collected through a controlled wiki experiment as well as data proposed to be collected in a case study in Second Life. The EMIL simulator is being designed to support a range of alternative assumptions and so should allow us to test alternative hypotheses and contribute to our understanding of this increasingly significant phenomena.

References

Almeida, R. B., Mozafari, B., & Cho, J. (2007). On the Evolution of Wikipedia, *International Conference on Weblogs and Social Media* Boulder Colorado.

Becker, B., & Mark, G. (1997). Constructing Social Systems through Computer Mediated Communication. Sankt Augustin, Germany: German National Research Center for Information Technology.

Berger, P. L., & Luckman, T. (1972). *The Social Construction of Reality*: Penguin.

Bicchieri, C. (2006). *The Grammar of Society*. Cambridge: Cambridge University Press.

Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia, *GROUP 05*. Sanibel Island Florida USA.

Ciffolilli, A. (2007). Phantom Authority, self-selective recruitement and retention of members in virtual communities: The case of Wikipedia, *FirstMonday* (Vol. 8).

Coase, R. H. (1993). The Nature of The Firm. In O. E. Williamson & S. G. Winter (Eds.), *The Nature of the Firm: Origins, Evolution and Development*. N.Y.: Oxford University Press.

Coase, R. H. (1995). *Essays on Economics and Economists*. USA: University of Chicago Press.

Davis, J. H., Schoorman, D., & Donaldson, L. (1997). Towards a Stewardship Theory of Management. *Academy of Management Review, 22*(1), 20-47.

Davis, J. H., Schoorman, D. F., & Donaldson, L. (1997). Towards a Stewardship Theory of Management. *The academy of management review, 22*(1), 20-47.

Demil, B., & Lecocq, X. (2003). Neither market or hierarchy or network: The emerging bazaar governance (pp. 36): Université Lille/Institut d'Administration des Entreprises.

Donaldson, L., & Davis, J. H. (1991). Stewardship Theory or Agency Theory: CEO Governance and Shareholder Returns. *Australian Journal of Management, 16*(1), 49-65.

Donath, J. S. (1998). Identity and deception in the virtual community. In P. Kollock & M. Smith (Eds.), *Communities in Cyberspace*. London: Routledge.

Gibbs, J. P. (1981). *Norms, Deviance and social control: Conceptual matters*. New York: Elsevier.

Giles, J. (2005). Internet Encyclopaedias go head to head, *Nature*.

Habermas, J. (1976). Some Distinctions in Universal Pragmatics: A working paper. *Theory and Society, 3*(2), 12.

Jones, C., Hesterly, W. S., & Borgatti, S. P. (1997). A General Theory of Network Governance: Exchange Conditions and Social Mechanisms. *Academy of Management review, 22*(4), 911-945.

Jones, C., Hesterly, W. S., & Borgatti, S. P. (1997). A General Theory of Network Governance: Exchange Conditions and Social Mechanisms. *Academy of Management Review 22*(4), 911-945.

Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). *He Says, She says: Conflict and coordination in Wikipedia.* Paper presented at the Computer/Human Interaction 2007, San Jose USA.

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology, 56*, 365-392.

Rossi, M. A. (April, 2004). Decoding the "Free/Open Source (F/OSS) Puzzle" - a Survey ofTheretical and Empirical Contributions (pp. 42): University of Sienna.

Sanger, L. (2005). The Early History of Nupedia and Wikipedia: A Memoir, *Slashdot*.

Sanger, L. (2006). The Nupedia myth, *ZDNet.com*.

Sanger, L. (2007). The Early History of Nupedia and Wikipedia: A Memoir, *Slashdot*.

Searle, J. R. (1969). *Speech Act: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Stiles, W. B. (1992). *Describing Talk: A Taxonomy of Verbal Response Modes*: Sage.

Stvilia, B., Twidale, M. B., Gasser, L., & Smith, L. C. (2005). *Information Quality Discussions in Wikipedia*. Illinois: Graduate School of Library and Information Science.

Therborn, G. (2002). Back to Norms! On the Scope and Dynamics of Norms and Normative Action. *Current Sociology, 50*(6), 17.

Williamson, O. E. (1996). *The Mechanisms of Governance*. New York: Oxford University Press.

Williamson, O. E., & Winter, S. G. (1993). *The Nature of The Firm: Origins, Evolution and Development*. New York: Oxford University Press.