

2004

Credibility and authority on internet message-boards

Ryan Goudelocke

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses



Part of the [Mass Communication Commons](#)

Recommended Citation

Goudelocke, Ryan, "Credibility and authority on internet message-boards" (2004). *LSU Master's Theses*. 3190.

https://digitalcommons.lsu.edu/gradschool_theses/3190

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

CREDIBILITY AND AUTHORITY ON INTERNET MESSAGE-BOARDS

A Thesis

Submitted to the Graduate Faculty of
Louisiana State University and Agricultural
and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Mass Communication
in
Theanship School of Mass Communication

by

Ryan Goudelocke

B.A., Oglethorpe University, 2000

August 2004

Acknowledgments are due

Dean John M. Hamilton

Assistant Professor Emily Erickson

and especially

Associate Professor David Kurpius

for their able assistance.

Additional acknowledgements are due

Josh Noel

for his help in the execution of this research.

Table of Contents

Acknowledgements.....	ii
Abstract.....	iv
Introduction.....	1
Present Goals.....	4
Literature Review.....	7
Historical Grounding.....	7
Abstract Considerations.....	10
Models of CMC - Limitations Model.....	11
Models of CMC - Opportunities Model.....	16
The Role of Lexical Tokens in CMC.....	21
Methodology.....	32
Locating Appropriate Communities.....	33
Units of Analysis.....	34
Operationalizing.....	37
Parameters.....	40
Coder Reliability.....	43
Analysis - Thread Depth.....	45
Analysis - Support Response.....	47
Discussion.....	51
Why These Tokens?	59
References.....	63
Appendix: Raw Data With Source Notes.....	68
Vita.....	74

Abstract

This research aimed to provide some proof or refutation of the hypothesis that online communities develop specialized vocabularies, often technical jargon, and use elements of those vocabularies, here labeled “tokens,” to ascribe credibility and/or authority to other posters. The literature from a variety of communications fields relating to this topic was summarized as a progression from an early “limitations” model of computer-mediated communication (CMC) to a later “opportunities” model. The drawbacks of current research were outlined and some new paths were sketched, including the methodology employed here.

Several discussions from different Web sites, each containing hundreds of posts, were tabulated and analyzed for the effects of inclusion of anecdotally-chosen “token” posts. Gauging authority and credibility as attention paid token posts and positive reaction to token posts, respectively, no correlation was found between token posts and attention paid them. One of three discussions showed a strong correlation between token posts and positive reaction, while two other discussions analyzed yielded results short of statistical significance. Suggestions were made regarding further work in this expanding field.

Introduction

As long as computers have been connecting people, researchers from a variety of fields have been puzzling out what changes might overtake all factors involved in electronic exchange of information: how will technologists adapt machines to suit human needs? How will society respond to increasingly quick access to massive flows of information? And how will interpersonal communications change when conducted over "the wire," sight unseen?

Answering those questions means asking them in research settings - something that has been happening only in the very recent past. Computer-mediated communication (CMC) has rapidly become a mass medium whose modes of transmission developed "under the radar" of communications workers. Features of online communication as simple as adaptations of spoken and written English to CMC have until recently been poorly understood. Only within the last several years, in fact, has communications research moved away from a narrow understanding of CMC as a strictly technological development and accepted online "culture" as such.

Accompanying that acceptance has been a marked shift away from methodologically rigorous experiments, which concentrated on the computer hardware itself essentially as a newfangled telephone or bulletin board, toward

more holistic overviews of online communities. But while the development - the explosion, really - and nature of those communities, which coalesce around Web sites, newsgroups and other gathering places, has attracted plenty of attention, the structural habits of group communication remain murky. Understanding online interpersonal exchange requires widening the scope of research into the textual habits of communicators, the lexicons of communities formed, and persuasive strategies adapted to online use. An essential part of that textual analysis is studying how members of online communities differentiate between texts produced by other members and those produced by outsiders, and how they discriminate between the two. Current research tends in that direction, and this content analysis should help shed light on specialized textual signals used by group members to filter information.

Accepting those signals, here called ‘tokens,’ as a useful area of research first requires definite proof that they exist and that they perform those functions hinted at by previous work: community organization, credibility enhancement and filtering online ‘noise,’ to suggest a few.

Originating in a strongly organizational, task-oriented methodology, early CMC laboratory research directly compared transmission of (sometimes identical) information in face-to-face (FtF) communication and via computer. Research questions often took the form of efficiency surveys, number of typed characters

used per communication, and the terseness or verbosity of particular subjects' typing. Working from assumptions that CMC would remain technologically and experimentally limited, theorists employed what I have labeled the "limitations" model to describe online behavior.

Early CMC research was generally conducted in artificial settings and employed quantitative methods - e.g. Ferrara et al's 1991 simulation of travel-agent interaction (see below) - in substantial contrast to later work in what I have called the "opportunities" model. While moving research into the existing world of CMC, rather than recreating it in the lab, workers also espoused ethnocultural and rhetorical bases and largely abandoned statistical methods in favor of case study and informal analysis.

Later CMC research generally repudiated the old ideas of limitation as artifacts of rudimentary technology, and broadened during the 1990s to embrace the potential development of new communications methods "over the wire" rather than concentrating on their limitations. While early work had gauged CMC on its users' capability for accomplishing tasks or transmitting information that might as easily have been spoken face-to-face (in short, its capacity to save a telephone call or a walk down the hall), new ideas of relationship and group formation widened the cultural-development possibilities that a many-speakers/many-listeners medium might hold.

Present Goals

This paralleled a tendency for CMC researchers to delve into socio-psychology, exploring phenomena like self- and group-formation, self-disclosure, even Baudrillardian and Nishidan questions like the significance of the mask and self-construction of identity. Those sometimes esoteric issues represent the extreme of high-level thinking about self- and community development, and relate to CMC only incidentally. Some workers closer to the ground, as it were, have noted along the way - especially those concerned with linguistics and rhetoric - the common phenomenon of group-specific lexical development, but the adoption of common jargon, abbreviations, even completely obscure symbology within particular groups has been treated merely as convenient placement of signposts in cyberspace.

That identification of tokens as simple hallmarks of mature online communities, with bare hints of more below the surface, is about as far as lexical analysis has proceeded in CMC. But tokens, by their operative definition as community identifiers and credibility enhancers, are easily susceptible to traditional investigation. Identifying what tokens are in use by any particular CMC community requires recourse to classic CMC methods of text coding and quantitative token analysis. Such analysis can take advantage of the text-only nature of traditional CMC and pick through online text messages with an aim to

finding out whether they constitute a credibility system: whether group communications contain embedded signals that help group members decide what messages should be given attention and taken as legitimate.

This adoption of content analysis does not reject the opportunities model in favor of limitations; indeed, the opportunities model may need to adopt more rigorous methods precisely because it ascribes such wide-ranging cultural and normative powers to CMC. Those researchers who have applied quantitative methods to, for example, Usenet discussions have employed traditional rhetorical criteria in their analysis.

This analysis is a quantitative study of communities and their use of textual signals - Web site message-board discussions, in this case - to obtain as representative as possible a sample of those communities' use of group-specific tokens and community reactions to their use. In practical terms this means identifying a set of tokens in use by a particular community, parsing of a sample of discussion text which contains these tokens as well as various other content, and analysis of reactions to token use to determine whether token use enhances or diminishes credibility.

Following is a review of literature surrounding CMC, especially that describing the development and analysis of linguistic/textual cues embedded within electronic discourse. CMC has grown into a multi-disciplinary field,

including sociolinguistics, rhetoric, semiotics and education, in addition to communication. In its early stages, however, researchers employed simplistic theoretical reduction of CMC effects to their most salient factor - the computer. This focus on the machine would go more or less out the window within just a few years, as technology advanced and CMC became less the exclusive province of government and academicians and more a people's playground. In today's wired society, of course, a research discipline untouched by CMC would be as rare as, for example, a hobbyist club or special-interest group without a Web site.

Literature Review

Historical Grounding

One of the earliest questions raised about CMC was its utility as a medium allowing the development of community, subject to various definitions of that word. A touchstone for this early work was Anderson's (1983) work on "imagined communities." In the tradition of Max Weber and Hannah Arendt, Anderson's inquiry concerned the origins and spread of nationalism, but his insights about self-organizing groups have been widely applied to the ways like-minded people congregate and construct communal spaces on the Internet.

Imagined Communities proposed a variety of criteria for defining a (national) community, a couple of which ran directly counter to the eventual development of online groups: first, necessary unfamiliarity with the vast majority of the group's members. In Anderson's formulation this was enforced by the sheer scale of national communities, while online groups form within precisely that personal familiarity; and second, necessary self-definition in terms of opposition to existing groups - while online groups may spring into life out of frustrations with a progenitor, they rarely define themselves in oppositional terms.

The question of defining "community" in CMC terms has certainly not gone away, though the terms of debate have changed somewhat. In purely prescriptive

terms, some writers now concern themselves with community development as an idea of changing processes (Miller, 1996). Analyses of information infrastructure, for example, advocate from a progressivist viewpoint social benefits like widespread information access, wired schools and online multiculturalism.

One early application of traditional mass communication theory to media-driven community development is acculturation theory (Bennett, 1982), an explanation of media effects on cultural norms - faith and trust in reality reflected in mass media, for example - as opposed to later recognition of media (including CMC) influence in development and origination of new and changed culture.

A major contributor to CMC research in its early days was a grounding literature in sociolinguistics that would be widely applied, in concepts and terminology, to questions of information transmission and community development via computer. Use of the terms "register," "genre" and "dialect" became more common in CMC research as linguistic and textual variations attracted more attention as potential signals of group-specific development online. One linguistic survey used those terms very early (Hudson, 1980) along with a prescient discussion of "speech community" as a set of shared attitudes, not simply a common medium or channel within which information is exchanged.

Alongside these questions of how to describe what online speakers are saying, CMC researchers have needed to draw on rhetorical analysis to understand

why they are saying it. Terms like "recommender systems" and "credibility systems" have been used alongside traditional rhetorical terms like "legitimacy" and "authority" to describe the interplay among online group members and group outsiders. Traditional rhetorical concepts like persuasion have been applied to CMC at the message level, but linkage of textual manipulation as rhetorical strategy has not occurred (O'Keefe, 1990). This analysis argues for the exploration of that link between the wording and lettering of text use online and the writer/speaker's rhetorical credibility with her audience.

As will be noted below, task-oriented and functional interpretations of CMC dominated early research, and the influence of linguistic disparities among communities continues to arouse debate, inside CMC research and out. One study of "information and referral" networks, without explicit reference to CMC, decried the absence of standardized terminology in social welfare work (Levinson, 1988). Social scientists are, of course, intimately concerned with employing the same terms, with the same meanings, as collaborators and colleagues; operationalizing words with dissimilar meanings contributes to confusion. Similarly, CMC work has begun to take more notice of textual variations, and the use of group-specific terms and jargon has become more important in defining and studying individual online communities (Crystal, 2001).

Abstract Considerations

Alongside questions of community formation in CMC runs a more abstract exploration of the psycho-cultural effects of transformative media like the Internet. The chronically overused word "cyberspace," for example, is a product of William Gibson's seminal 1984 novel *Neuromancer*, a post-apocalyptic rhapsody of twisted electronic media commercialized and manipulated beyond recognition. In Gibson's novel, as in Ridley Scott's 1982 "cyberpunk" film *Blade Runner*, mankind lives in opposition to an inescapable technological infrastructure that has swallowed him whole.

Seeds of similar ideas are present in early applications of traditional mass communication theory to technological media (Curran et al, 1982) as well as more optimistic, expansive reactions to pervasive electronic media (McLuhan, 1964). Their opposition might well be seen as one of expansion versus alienation, paradigms of the computer and the network as liberators, empowering individuals and communities, or as jailers destroying them (Meyrowitz, 1985).

As time has passed and people (and researchers) have become more familiar, more comfortable and more "socialized" (Meyrowitz's usage) with the uses and limitations of information technology, a more integrative viewpoint has become common to such abstract ideas of community, a point of view Heim (1993) called "the computer as component" of larger cultural change. That integrative viewpoint

gradually strengthened in online communications studies during the 1990s, as technologically-limited research methods became a point of contention.

Models of CMC - Limitations Model

Early CMC research generally arose out an attempt to quantify the value of computers and networks for organizational tasks, and thus took the form of short, goal-oriented sessions in laboratory settings, experiments limited to the comparison of (sometimes identical) CMC and face-to-face exchanges.

Findings from this line of research have generally emphasized the social disadvantages of computer-mediated communication, therefore implying that highly developed, positive personal relationships should occur infrequently in online settings.... Relational cues emanating from the physical context are missing, as are nonverbal cues regarding vocal qualities, bodily movement, facial expressions, and physical appearance. CMC is thus judged to have a narrower "bandwidth" and less information than FtF communication (Parks and Floyd, 1996)

Valuation of CMC in a task-efficiency sense remains current in some contexts which prize its ease of distance and network transmission and are less concerned with mechanics (Levinson, 1988). For example, online communication as an educational tool must meet a variety of goals in order to appeal to its users, both educators and students, and surpass traditional learning networks in usefulness (Hoadley and Pea, 2002). Some research has directly compared developed, pre-existing virtual networks to traditional social ties to tease out which

can be more valuable at disseminating learning materials (Haythornthwaite, 2002).

But task-oriented, organizational CMC proceeded along evaluative lines that mandated comparison with FtF communications and "traditional" channels like the spoken and printed word, telephone, radio, and television. In the main, the idea of CMC as an outlet for complex group formation and culture development suffered from the comparison. Some later descriptions of online self-construction and individual formation stuck to this viewpoint of CMC alienation, notably in psychological terms like "fragmentation of self" (Kolko and Reid, 1998). Their overview of CMC "failures" describes problems that arise from online personae yielding, essentially, less than the sum of their parts: "... multiple instances of the self may allow the individual projecting them to experience a greater diversity of himself or herself than would otherwise be the case, but each single instance operates on a very limited psychological and social plane" (218-219).

These tales of alienation reflect a deep-rooted conception of CMC as a dividing and fragmenting influence on self- and community development. Despite potential experimental deficiencies (see below), early work understandably produced some of the same aspects of communication as noted by later researchers: chief among them privacy concerns, ease of anonymity and potentially lower accountability (Elgesem, 1996). While CMC in general seems to bring such issues to the fore, early interpretation of their effects differed significantly from

most later work: experimentation described loss of empathy, hypersensitivity, disinhibition, anti-social behavior and de-individuation producing lowered personal involvement and investment (Kiesler et al, 1984; Lea and Spears, 1991). Interestingly, exactly the same effects were re-interpreted, with sometimes starkly different conclusions, as researchers took their studies out of the laboratory and into "the wild" and began looking for the role of anonymity, disinhibition and self-construction in cultural formation and development (Danet, 1998; Lyon, 1997; Wiszniewski and Coyne, 2002).

The recognition that CMC, in strict comparison to FtF and traditional channels, is devoid of physical context (Shaw, 1997), combined with descriptions of anti-social behaviors like "flaming" (Parks and Floyd, 1996) produced what later researchers variously termed the "cues-filtered out" (Baym, 1995), "filtration of cues" (Parks and Floyd, 1996) or "filtration of appearances" (Rheingold, 1998) model, all of which in retrospect might be lumped as a "limitations" model to emphasize their differences with more recent work.

Interestingly, even while researchers viewed the textual limitations of CMC as inherently limiting beneath the capability of traditional channels, some already were using language which hinted at the revaluation of textual cues *qua* new form of communicative "signal": "[CMC suffers from] the absence of [traditional cues] to enable the intended decoding of a typed message" (Shaw, 1997). In other words,

the vagaries of textual analysis dictate that a speaker can't be sure her speech will be reconstructed by readers with the shades of meaning, emphasis and tone she tried to convey. Use of the word "decoding" points to an important understanding of CMC texts as potentially encoded messages (Raymond, 2000) or as texts whose meaning and relevance may supplant or replace the traditional cues technology has filtered away.

Even within the conception of CMC under the limitations model arose notice of "paralinguistic" features evident in experimentation as goal-completion strategies (Wilkins, 1991), among them "lexical repetition," or the adoption of certain words as key landmarks within a chat discussion in order to maintain multiple independent topics amid chatroom chaos. The absence of traditional cues in interpersonal CMC also gave rise to emotional signifiers inserted within texts. Emoticons like the 'smiley,' for example, arose quite early in CMC's history and spread rapidly, evolving many forms along the way.

Other "paralinguistic" channels formed in response to the stifling effect of CMC filtering, enabling communicators to manipulate text elements to convey not just emotions but emphasis and argument structure with, for example, **boldface text**, *italics* and CAPITAL LETTERS - the last the online equivalent of shouting. Other textual means of emphasis included the use of multiple vowels (e.g. "that test was sooooooooo difficult") and the use of asterisks or underscores to **emphasize** a

point. One researcher (Murray, 1991), uniquely, included with an early discussion of the "smiley" emoticon an example that requires a head rotation to the *right*, a "left-handed smiley", written as "(:" or "(-:". Smileys and other emoticons, because they generally terminate a phrase or sentence, usually require a head turn to the *left*, written as ":)" or ":-)."

The influence of linguistic usage on ideas of CMC was noted by an attempt to categorize one-to-one communication via keyboard as "interactive written discourse" (IWD) (Ferrara et al, 1991). This model labeled computer communications an "emergent register," in the sociolinguistic sense of register as a "type of language employed in specific situations" as opposed to genre, a "type of language employed by a specific speaker" (Hudson, 1980; Geertz, 1973). In Ferrara et al's experiment, subjects conversed with a "human-assisted system" travel agency, composed, unbeknownst to them, of a single human (one of the experimenters). The design tested "omission of subject pronouns, copulas and articles" and gauged the degree of grammatical formality employed to reach conclusions about "computer talk," an emergent register the authors compared with, e.g., Baby Talk, Foreigner Talk and Bureaucratic Talk.

Ferrara et al's experiment typifies the experimental design employed by limitations model research, design that has been criticized for artificiality of setting, limited time allotted to group formation and serious technological

limitations. In all fairness, the absence of a large population with regular access to CMC technology and the state of available consumer technology until the mid-1990s played a major role in limiting the depth of CMC exploration. Indeed, the early 1990s marked the first widespread, non-technical and cultural-specific, rather than task-specific, use and study of CMC technology outside artificial settings (Baym, 1998). It's possible that researchers working before the popularization of Internet Relay Chat (IRC) and Usenet newsgroups found little community cultural development simply because there was little or nothing of the kind happening.

Models of CMC - Opportunities Model

Possibly the most influential restatement of CMC's potential and challenge to the limitations model was Howard Rheingold's 1993 book on virtual communities. His essential insight was a distancing from the computer-dominated perspective of early CMC work to a more grounded consideration of how people interact online in natural settings: he asserted that virtual community-building simply requires the same familiarity and common interests and dispositions that give rise to real life (RL) groups. Significantly, he faulted then-current research for attempting to construct artificial CMC groups in unrealistically short timeframes and in artificial experimental settings.

Quoting J.C.R. Licklider in the April 1968 issue of *International Science and Technology*, Rheingold agrees that CMC groups "will be communities not of

common location, but of common interest" (quoted in Rheingold, 1998): a simple assertion but one that had so far escaped communications theorists, who reached conclusions drawn largely on experiments in laboratory settings.

Important insights in Rheingold's work also include the prediction that online communities develop novel cues (though he does not use that term) as they grow; that "sifting the key data" from the flood of information available online will be an important function of CMC groups (see Crystal below); and that online communities will exert normative influences on their members and on the forms in which they communicate - including linguistic and textual forms.

Compared to the theoretical work he summarized, Rheingold's challenge casts something of a wide net. He is concerned not with abstract notions of the mechanics of CMC - indeed, not with CMC specifically at all - but with the influence of online culture on aspects of everyday life. Lamenting the overuse, in academia and popular media, of the phrase "virtual community" and the ever-present buzzword hype surrounding "virtual reality" and similar concepts, Wilbur (1997) argued that online community should be conceived in open-ended terms:

Perhaps multiple, contradictory definitions look considerably less useful than, for example, Rheingold's fairly elegant, singular attempt. However, the point of all this memetic dissection is not to better fit the words "virtual community" to some known social reality. Instead, we are at a point in our researches into Internet culture where it is particularly important not to force old conceptions - such as that of the

mythical frontier, for example - on the new phenomenon of decentralized networks of multi-tasking, time-sharing machines and human-machine interfaces. We do not know very much about Internet culture, so perhaps the best definitions are multi-bladed, critical Swiss Army knives (15)

In this broad view of Internet studies, CMC becomes a blade best suited to dissecting communications mechanics, i.e. how information is presented and transmitted to exert influence on a broader online culture. Part of that view, as CMC research matured, meant acknowledging the superior relevance of cultural ethnography in community formation over the old organizational, goals-oriented focus (Giese, 1998).

Structural considerations in CMC also moved away from technological limitations and experimental design to the observed structure of communities revolving around shared culture (Mitra, 1997). Mitra observes use of paralinguistic features and textual image-description as tools to construct virtual spaces; in one case, a community of Indian expatriates sharing family stories and contemporary political discussion intermixed with reminiscences of home. This sort of qualitative analysis of existing CMC communities, often observational study of newsgroup and discussion board archives, offers a valuable way to watch alternative textual cues develop (Baym, 1995).

One insight from Nancy Baym's observations of a soap-opera fan newsgroup on Usenet is the growing self-awareness she described on the part of group

members: "Users are aware that their cultures have group-specific forms of expression and take active roles in the codification of those expressions" (152).

Mitra had in mind attempts to set a visual scene in text, however, and Baym's "group-specific forms of expression" had more to do with norms of etiquette than with textual tokens having legitimacy or credibility roles. She simply described members using "affirmative" or "supportive" language, without further specifics, as a mechanism for maintaining a friendly group atmosphere.

The highest-level research into online communities, be it from a "structural" or "cultural" or some other viewpoint, generally ascribe less importance to textual manipulation than to composition (in Mitra's case) and emotional support (in Baym's).

A large body of later CMC research concerned with self- and group construction online performed an about-face repudiation of earlier conceptions, but there was little recourse to textual analysis. Rather, it was identifying freedom of self-disclosure and lack of "social gatekeeping" (McKenna et al, 2002) as an important component to relationship formation, and valuing anonymity in group formation, that entailed a direct denial of the limitations model. As noted above, previous work had condemned anonymity and disinhibiting behavior as counter-productive (Lyon, 1997). Reference to the ability of "Netizens" to assume or shed anonymity at will as liberation, the self as a "project to be constructed" and a

description of the "mask" as essential to self-construction complete the adoption of a cultural-development "opportunities" model of CMC (e.g. Danet, 1998) - a model that relies on informal impressions and generally eschews hard textual analysis.

In keeping with Rheingold's predictions, according to opportunities formulations, online groups develop best when members experience disinhibition - the same factor that gave rise to anti-social behavior in controlled settings - as a way to experience new viewpoints to be shared with community members. In essence, opportunities theorists embrace the filtered cues of CMC as a freeing mechanism (Wiszniewski and Coyne, 2002) instead of condemning them as communications failure.

But the importance of textual cues, especially the role they play in replacing or expanding on traditional cues, is evident in identity studies that blend rhetorical considerations with specific reference to linguistics. Examination of whether CMC gender roles reflect RL prejudices and preconceptions, for example, required analysis of what text cues signified condescension or patronization (Herring, 1996). Even more abstract discussion of gender roles online, and gender switching by Netizens, needs a typology of what textual cues are masculine and which feminine, and more broadly which cues indicate approval and which opprobrium (Davidson and Schofield, 2002).

While the opportunities model allowed researchers to accept a much larger

role for online communities and the complexity of their interactions, rigorous analysis at the text level became deprecated in favor of more holistic conceptions. With some exceptions, there has been little interest in the role low-level text manipulation may play in group self-identification and development.

The Role of Lexical Tokens in CMC

Computer-mediated communication thus described is, obviously, a topic at which several research fields converge. Rhetorical questions, persuasion and advocacy and recommender systems, for example, can be conceived as the interplay of textual cues developed and used online (O'Keefe, 1990; Gurak, 1997; Avery and Zeckhauser, 1997).

On a more abstract level, CMC texts are analyzed from a standpoint of their effects on historical language composition in what might be (whimsically) called the 'semiotic superhighway' (Shank and Cunningham, 1996). A thought experiment introducing Shank and Cunningham's discussion envisions a modern-day Descartes: the French philosopher downloads an e-text of his own *Discourse on Method*, reads perhaps a paragraph or two, then sends a hastily-composed "flame" email based on what little he has read. Hilarity does not ensue; this warning of the tendency of electronic media to shorten attention and reduce complex gradations of thought as simply as possible dates at least to McLuhan (1964), and similar ideas about radio and (especially) television preceded him.

Though CMC includes these strands, and others, concerning abstract theorizing, the field remains somewhat grounded in analysis and manipulation of the text as the center of group interaction. Thus some literature does address language manipulation as hallmark or signifier of particular online communities: how textual cues aid in rhetorical establishment of legitimacy (Connery, 1997) and the role of emoticons and syntactical cues in text composition (Murray, 1991). This study of linguistic tokens as signals of online communities does not, however, extend to instrumental roles tokens may also have.

There also exist a variety of compilations of CMC-specific abbreviations like 'YMMV' (Your Mileage May Vary) or 'IMHO' (In My Humble Opinion) (Connery, 1997). Many others, like 'ROFL' (Rolling On the Floor Laughing) and the ubiquitous 'LOL' (Laughing Out Loud) have sprung up, along with some cheeky mutations: 'IMNSHO' (In My Not So Humble Opinion) and 'ROFLMAO' (Rolling On the Floor Laughing My Ass Off), for example.

However, little attention has been paid to the difference between CMC-specific and group-specific cues, and what if any internal uses group-specific cues serve the group itself. Generally the existence of community-specific lexical tokens is acknowledged (Baym, 1998), though that acknowledgment is limited to the rudiments above: emoticons and abbreviations, and (rarely) passing mentions of group-specific jargon.

Use of the term "token" has been narrowly employed in past CMC research, specifically in linguistic analyses of the economy of language in online speech. The linguistic concept of "type-to-token" ratio, applied to an entire text, purports to measure the frequency of "informational" words, or tokens, versus the frequency of "instrumental" or functional words. Similar cross-sections of a text may be derived by imposing one or another token typologies to words used in a text. This linguistic research has much in common with IWD work; indeed, CMC is lumped into an "electronic language" exactly akin to "computer talk" (Collot and Belmore, 1996). However, type-to-token ratio work within CMC has focused on traditional rhetorical strategies (e.g. modality, personal reference, semantics) and avoided the question of typological variations within particular online communities, much less used group-specific linguistic tokens as identifiers or examined them in rhetorical terms (Yates, 1996).

Use of the word "token" for a package of characters which, taken together, hold some linguistic meaning is a useful prior step to understanding "token" as it is used in this research: as a signifier, a sometimes-cryptic piece of encoded information that requires specialized knowledge to "unpack," like some medieval literary allusion. The essence of tokens as enhancers of credibility and authority is in their dual nature: to the world outside an online community, a token has the bare minimum, linguistic meaning that IWD work understood simply as a piece of a

sentence. From within, however, community members see tokens as a transparent shorthand for hidden meaning.

As the absence of traditional cues influenced limitations theorists to conclude that CMC simply lacked alternative channels, so the employment of group-specific tokens as more than passive indicators that a community simply exists apparently has escaped contemporary opportunities researchers. Researchers have acknowledged that group members use peculiar words or phrases, and that those phrases often mark internal group development, but individual use of those tokens has not been studied as rhetorical strategies in their own right.

One of the few examples of group-specific tokens Baym (1998) relates from her observations of the soap-opera newsgroup "r.a.t.s" (rec.arts.tv.soaps) is common nicknames for soaps characters: When "Natalie" was involved in a fatal traffic accident, her common nickname on the newsgroup morphed (by popular usage) from "Nat" to "Splat." She also mentions a r.a.t.s-specific abbreviation: IOAS, or 'It's Only A Soap.' Baym also (1998) refers to an unpublished 1995 dissertation by Cherny (unavailable) that references community-specific emoticons, abbreviations and jargon as language "registers" in the sociolinguistic sense. Even while pointing out r.a.t.s-specific tokens, however, Baym confines her rhetorical analysis to traditional means.

A definitive treatment of the development of group-specific tokens - the

term attempts to cover emoticons, jargon, in-jokes, even technical information - is available in David Crystal's (2001) encyclopedic survey of English usage on the Internet. Purposely avoiding both 'CMC' (because it emphasizes the computer) and the various terms 'speech communities,' 'register,' 'genre,' and 'discourse type' (he adopts 'variety' instead), Crystal advocates a linguistic approach to online communities that equates them with RL groups, only more dependent on textual nuance for conveyance of meaning.

Despite Crystal's authoritative exploration of what CMC jargon is in use, he virtually omits discussion of the instrumental function of tokens. He does offer an enticingly brief mention that the jargon adopted by group members may serve an information-sifting role, but he does not make the leap into assigning functional filtering significance to group speech.

Crystal treats 'hacker' jargon as the foundation for what he calls 'Netspeak,' even for people who are not technical professionals. Crystal notes that the term 'hacker' is itself a jargon word; in popular usage it has taken on a negative, even illegal connotation, while in its original sense - which survives in the hacker community - it means anyone who programs computers professionally or recreationally. True hackers call people who engage in illegal computer activity, such as breaking into systems illegally, 'crackers' or 'black hats.'

Crystal's only nod to the normative function of group-specific tokens mirrors

Baym's (1998): "... [M]embers accommodate to each other.... [T]heir contributions progressively develop a shared linguistic character - the equivalent of a local dialect or accent. Everyone comes to use certain types of grammatical construction, slang, jargon or abbreviations" (Crystal, 147). Anyone who has traveled in an unfamiliar region knows how quickly accent, even in a shared language, marks one as an outsider.

Again, use of such group-specific tokens is understood in the literature simply as a signal of group existence, not in an instrumental sense of specific intentions - intentions that may include concealing meaning from all but those "in the know" or actively excluding non-members from discussion. Among members of highly developed online communities, the use of particular group-specific tokens not only identifies members (Crystal's use of the word "cognoscenti" emphasizes shared understanding, not just shared attributes like accent or dialect) but serves as a powerful credibility and authority strategy - leading other members to concentrate on in-group speech and disregard other speakers.

When the question of CMC credibility strategy has arisen, theorists have had recourse to traditional rhetorical considerations like sufficiency and speaker legitimacy (Liu, 1997). From this viewpoint, CMC texts attain rhetorical strength in the usual ways: ingratiation to the reader, sufficient attention paid various reader concerns, assertion of the speaker as an authority or expert, and so on.

Though Baym's (1996, 1998) are qualitative case studies, she does note the use of various empathetic textual strategies: usage of signifiers like "I know what you mean;" offering similar personal experiences; and addressing other speakers personally all illustrate the application of traditional rhetorical devices to CMC discussion. They are not usages of group-developed slang or jargon, sometimes with meanings apparent only to the "cognoscenti." Tokens as used in this analysis instead serve as iconic signifiers of meaning, and can be employed rhetorically to signify membership and authority.

Adopting Liu's (1997) typology, Galegher et al (1998) quantitatively analyzed rhetorical strategy of self-help newsgroup members according to 'appropriateness,' 'relevance' and 'sufficiency.' These strategies included "using catchy headlines in the subject line of the header;" "signaling membership by behavioral statements" like descriptions of one's symptoms or interactions with medical authorities; and "asserting membership by self-identification" of how long the person has been reading the newsgroup. All of these, of course, are rhetorical strategies translated directly from traditional channels. Unsurprisingly, given that two co-authors (Sproull and Kiesler) both published significant research on the limitations model, Galegher et al. conclude that what is old is new again:

What may be most interesting about our findings is that these discourse patterns appear without any of the props generally

considered important in face-to-face conversation and without meeting the standards generally associated with carefully thought-out prose. The similarity of electronic discourse to more traditional forms of discourse is a testament to the flexibility of people's ability to convey their personalities and personal problems as well as their points of view in different rhetorical situations and technological environments (Galegher et al., 1998)

The brand-new formation of group-specific lexical tokens, in contrast, arises where rhetoric and online communications meet. To the extent that CMC researchers have considered them, they have been assigned the simple role of replacement cues for the traditional cues filtered away. Their true role may be much more significant, as influential hacker Eric S. Raymond declared:

As usual with slang, the special vocabulary of hackers helps hold their culture together - it helps hackers recognize each other's places in the community and expresses shared values and experiences. Also as usual, not knowing the slang (or using it inappropriately) defines one as an outsider, a mundane, or (worst of all in hackish vocabulary) possibly even a suit. All human cultures use slang in this three-fold way - as a tool of communication, and of inclusion, and of exclusion (Raymond, 2000)

Raymond may be overreaching in generalizing about human cultures, but 'hackerdom' clearly is one far-flung community that has developed a subset language of symbols that "communicate shared experiences" and, just as important, serve community members in identifying those who belong - and those who do not.

Crystal (2001) comes close to explaining this exclusionary rhetorical function of group-specific tokens when he discusses the common online practice of "trolling," or knowingly posting ill-informed nonsense or offensive material in hopes of provoking a reaction. Responding to a troll ("feeding the trolls") indicates non-membership; "old hands will simply ignore it or - if they can be bothered - laconically send the response 'nice troll' to the originator, or YHBT (You Have Been Trolled) to the responder" (53).

Crystal's prescription for understanding the unique, ephemeral 'Netspeak' in use among different CMC groups is "grounded in systematic empirical observation, providing a representative corpus of material which would reflect the frequency with which Internet situations use and vary particular structures" (73).

There has been no such systematic study of token use, something close to Crystal's "structures," in any online community. Nor do there exist studies which examine whether analysis of group-specific tokens could be useful for more than simple identification of a group - whether in fact their selective use constitutes an emerging system of rhetorical strategy unique to the rapidly evolving registers of CMC.

Operationalizing "token" based on the foregoing literature necessarily involves some ambiguity; the word may include technical jargon, abbreviations, in-joke references, trivia, minor details from a television program or movie, or

other textual objects. Starting from such a broad theoretical base, there is a danger of overgeneralization about what constitutes a token and what does not. The simplest way to ensure reliability in the identification of tokens would be a typology encompassing all possible token categories.

However, development of such a typology is not feasible because of the vast disparity among potential types: movie quotes, snippets of computer programming code, snatches of poetry, acronyms for common phrases, song lyrics, dialogue from television shows, book titles, names of people obscure to the public but well-known to the community - these are only a few of the many possible signals group members can use to signal credibility and authority. The problem of categorization is even worse; even knowing the origins of every potential token (an impossible task), the way they are employed can be situation-specific. A single token might be used to convey disgust or gentle good humor, depending on the writer/speaker's state of mind. And of course the very function of tokens - to shield from the outside the attention and time of group members - militates against any outsider's uncovering of their meanings. Tokens are weapons aimed away, at incoming strangers; they are ramparts facing outward.

Thus tokens may not be categorized in as straightforward a manner as, for example, rhetorical credibility strategies have been. Their significance is bound up in recognition and shared understanding, not their intrinsic rhetorical activity. A

serviceable metaphor might be a random pass phrase chosen to govern entrance into an exclusive club; the token is a vehicle for recognition, not necessarily a functional "text" in itself.

Further complicating matters is the presumption that tokens are, by definition, group-specific, so that a phrase or code common to one online community may be taken for gibberish by members of another. Thus analysis for the same sample token from different sources greatly decreases the potential for a meaningful result. These problems lead to the assertion that tokens selected for analysis must be chosen anecdotally and within the scope of a single community, and studied within particular discussions to whose subject matter they are relevant.

This broad theoretical stroke at defining what tokens are and what purposes they serve demands limitation if they are to be identified for study, counted and examined. Conceptually, they may serve a number of roles in addition to enhancing credibility and authority: efficiency of communication, for example, as a form of shorthand either for language or concepts left out. They may filter speech, red-flagging misused tokens for a quick skim instead of a close read. In this study, despite the theoretically vast galaxy of tokens and their uses, a single hypothetical correlation between tokens and enhanced credibility/authority is at issue.

Methodology

Research questions: do Internet posters ascribe greater authority and credibility to other posters who use anecdotally-chosen tokens? The lowest-possible unit of analysis to determine reactions among posters is the single message; authority is here operationalized as greater attention paid, or more replying posts. Likewise, credibility is assessed by judging whether replies to posts are more or less positive in response to the presence of the token.

Hypothesis I: Posters are more likely to respond to posts containing tokens.

Hypothesis II: Posters react more positively to posts containing tokens.

A variety of factors limit which and how many communities and discussions may be examined: some Web sites, for example, do not archive discussions. Those that do may display comments in a flat chronological list, rendering the task of tracing response reactions within posts impossible. Still others may nest comments in a parent post-child post order, allowing responses to be traced, yet make simultaneous display of all comments within a discussion impractical.

For these and other logistical reasons, this content analysis necessarily is a limited look at the operation of tokens on a small scale: a few discussions within a few communities. In the course of this analysis, recognizable tokens were identified much more easily in the context of technological discussion; by no

accident, all three of the discussions analyzed were debates largely about technology. Anecdotally, at least, discussions about technology appear to lend themselves to token usage, perhaps because technology subjects often involve specialized terms or jargon easily seized upon for community identification (Raymond, 2000).

This is not to say that tokens can not be isolated in other contexts; indeed, several Web sites devoted to popular culture appeared promising initially but lacked the search features necessary to view hundreds of posts without difficult navigation. An example is Nohomers.net, a discussion site revolving around episodes of The Simpsons cartoons. Tokens that appear frequently there include the acronyms OFF, short for Our Favorite Family; and THOH, short for Treehouse of Horror (the series' annual Halloween episodes). Other examples of non-technical token use are so-called "obligatory" television show quotes, generally from niche programs like *The Simpsons*, *Dr. Who*, *Red Dwarf*, *Futurama* and others.

Locating Appropriate Communities

What treatment of token function exists in the literature above suggests that online communities need only form and participate in group discussion (one and the same process) to begin developing lexical tokens. Reasonable deduction implies that long-standing communities would develop more and more structured

tokens than newer ones, making established groups more valuable for investigation.

Any established online community with archived discussions, then, would be appropriate for study. While tokens may appear in chat text (Internet Relay Chat, America Online Instant Messenger, Microsoft Instant Messenger and so on), the one-to-one nature of those channels disqualifies them for "group" status. Although the people comprising an online community may certainly use those systems, there is no reliable way to determine what "group" means in that context.

Web site discussion boards and Usenet newsgroups, then, are most appropriate for analysis. Because newsgroups do not commonly archive their postings (though there are paid services that do), content is most easily retrieved from Web sites' archived discussions. Well-traveled sites with large numbers of posts per discussion will yield the most data. This content analysis examined discussions from two communities: Kuro5hin.org, a culture and technology Web site offering criticism of news and essays; and Slashdot.org, a technology-oriented Web site that includes news on civil liberties and privacy rights with a technological focus.

Units of Analysis

Within discussions isolated from those communities, the appropriate unit of content analysis is the "post:" the written expression of a single user, the post in

aggregate makes up the discussion and is the means by which online group members communicate. Many online communities allow members to investigate each other by searching past posts. In systems (depending on Web site design) where users originate discussions, the first original post defines the discussion that will follow; in systems where only administrators or moderators may commence discussions, original posts react to the administrators. In both cases, subsequent posters react to other users, forming branching chains of response that can include hundreds of posts.

In many online communities, including both Web sites included in this analysis, built-in moderation systems allow users themselves to judge the quality of each other's posts through a ratings system, then tailor their viewing preferences based on moderation ratings. For example, preferences settings can be used to ignore posts moderated down as irrelevant or uninformative while highlighting high-rated posts.

While a user-driven judging system might be a gold mine of data on community perceptions, serious problems confront drawing conclusions from such ratings: moderation generally is anonymous, decreasing accountability; moderation up or down may be for a variety of reasons, a single reason, or none, and those reasons may be omitted or manipulated as the moderator wishes; moderators may pay more or less attention to particular subjects depending on their interests; and

abuses of moderation may render posts invisible, depending on preferences settings, to other users regardless of their content. This content analysis ignored built-in moderation systems and newly analyzed all posts within the discussions under study.

The research question is in two parts: attention paid and credibility ascribed. Coding of the discussions was broken into two corresponding steps: measurement of thread depth with and without token appearance, followed by quantification of all child responses into value categories of response. Thread depth measurement simply tested whether original parent posts - those which responded to the discussion itself - generated more-numerous responses when the original parent contained the token under study. Thus in the first coding step, thread depths in each discussion were measured and averaged to determine if original parents containing the token presaged deeper threads than original parents without the token. Counting thread depth, obviously, is an objective procedure; the only source of confounding error is the choice of the token itself, which will be discussed below.

The second part of the analysis sought to grade posters' comments according to their reactions to parent posts with and without the token. Each response to a parent post was coded on a three-increment ordinal scale according to simple criteria: did the child post respond positively, negatively or neutrally to the parent?

In an earnest debate, judging responses by those three criteria might present more or less difficulty: more because responses would analyze and equivocate, smearing black-and-white response into shades of gray; less because posters would consistently respond to the parent, rather than typing nonsense or engaging in flights of fancy. In the course of coding the sample discussions, negative responses (disagreement, qualification or correction) were found to greatly outnumber positive (agreement or endorsement) ones, while ‘neutral’ responses generally outnumbered both. Labeled ‘neutral,’ this third category might easily be called ‘other,’ as it encompassed relevant digressions and equivocation as well as incoherent ramblings, nonsense, irrelevant profanity, flaming, trolling and other amusements of Web site denizens. Three categories were used because of the paucity of positive responses as well as the great number of ‘other’ responses; more granularity along ‘positive’ and ‘negative’ axes would probably have necessitated category aggregation in any event, and would have increased coding difficulty and time required.

Operationalizing

While the literature reviewed above does recognize the existence of lexical tokens as signifiers of group identification, there generally is little recognition of instrumental activity through credibility or information-sifting. Thus this content analysis both newly posits the existence of a communications tool, as it were, and

seeks to identify that tool's activity. This lack of recognized criteria for the existence of a token introduces a significant subjective element to the analysis: the choice of a token operative within a particular community and a particular discussion. The simple approach of positing a token and then analyzing for its effects admittedly exposes the analysis to the accusation of syllogism, but that analytical approach is unavoidable in the absence of a reliable token typology and the lack of empirical groundwork.

If, as several writers noted above suggest and the author strongly believes, instrumental tokens play important authority and credibility roles, the current lack of rigorous analytical work in the area will be remedied in time. It also should be noted that similarly anecdotal operationalization of value categories, like “positive” and “supportive” versus “negative” and “destructive,” are the norm for rhetoricians and semioticians. These counterarguments, it is acknowledged, do not negate the criticism of subjectivity. It is hoped that the review of the relevant literature, pointing the way toward this investigation, along with the analytical results may shed some light on token operation even if this content analysis does not, and could not, entirely settle the research question.

Without doubt the least defensible part of this analysis is the process of choosing a token and, after the numbers are tallied, ascribing to it the effects predicted. The same very broad theoretical understanding of tokens and their role

explained in the literature review above dictates uncertainty about rightly or wrongly singling out promising jargon, shorthand, acronyms, etc. for examination. But if the best that can be said about a particular token is that it is correlated with greater attention or positivity, the worst criticism that can be leveled is equally cautionary about other media of communication: confounding variables are many, and examination of a particular, anecdotally-chosen, token faces no more and no less trouble than any field study of oral, visual or written communication outside CMC.

As noted above, the elements of the research question regarding authority and credibility are operationalized by thread-depth measurement and response-value coding, respectively. Because of the broad nature of the question, rhetorical analysis in further depth was left alone; the vast number of ‘neutral’ or ‘other’ responses hinted that such analysis will be a difficult path. Simply put, there was an enormous amount of ‘noise’ present in the discussions analyzed, leaving potentially little grist for the mill of rhetorical analysis. Meta-investigation of whether certain communities, certain subjects within certain communities, or even particular community members employ various rhetorical responses to token use are well beyond the scope of this analysis, which attempts simply to determine whether tokens exert observable effects.

Parameters

To reduce avoidable error, discussions were used from Web sites which allow display of all posts to a discussion within a single, lengthy, scrollable Web browser page. Early attempts to navigate comments nested within comments by use of hyperlinks, the only method available at many sites, proved fraught with difficulty. Thread depths and value-response coding were noted by hand on paper, with simple series of numbers representing depth of threads and lists of symbols denoting value-response: ‘+’ for positive, ‘-’ for negative and ‘o’ for neutral. Tallies were checked twice with a calculator before data entry for statistical analysis.

Because different tokens were identified and studied for each discussion, the sampling unit (Riffe et al, 1998) was the discussion, or group of posts organized around a particular subject. For both Web sites under study, users are not directly free to initiate discussions willy-nilly; rather, site administrators choose from among user submissions which subjects will be highlighted. Neither site removes posts for any reason, meaning users are free, if they choose, to post about literally anything. Three discussions were analyzed; each was composed of hundreds of comments posted over several days (though the discussions were not contemporaneous). The three sampled discussions contained roughly 1,000, 600 and 300 posts, respectively. Each was active for a few days between late August

and early November, 2003. A control discussion, used for intercoder reliability, contained about 200 comments.

The potential for confounding variables is high, one of the most obvious reasons being the content of the post itself. The employment of standard rhetorical strategies, with or without tokens, may certainly confound data by influencing replying posters and inclining them to support the original post. Because the presence of a token is the sole independent variable, there is no possible control sample for analysis other than those posts which do not contain the token under study. Thus the control population must be the entirety of the discussion's responses.

Coding protocols included: for thread depth measurement, the original parent post was not counted; threads with no replies were counted at zero depth. Use of the token was counted in the subject header of posts, unless it appeared only in reply to a parent post using the token within the header. Similarly, use of the token in quoted material was not counted unless the poster also used the token in original material.

Despite the simplicity of a three-increment scale, and the consignment of irrelevant and incoherent material into the 'neutral' category, judging positive or negative response to parent posts sometimes proved difficult. There was no dearth of irony and sarcasm, nor of equivocating responses which praised or criticized

parent posts with qualification. With few exceptions, however, qualified agreement and qualified disagreement were coded in their respective categories in order to maximize useful data. There were few real shades of gray which defied classification as positive or negative; the few truly neutral, equivocal responses were swamped, in all three discussions (and the control discussion), by various species of troll, flamebait and irrelevance.

Identifiable self-replies (one poster responding to his or her own post) were counted as neutral, and in each of the few such cases encountered, consisted of clarification or addition to the foregoing post; neither site allows users to edit posted material. An important potential problem arises with the freedom to post anonymously, as both Web sites used in this analysis allow. Users are free to respond to their own posts anonymously, and presumably would do so positively, skewing the response analysis. However, token-using posters should not be expected to do so disproportionately more than non-token-using posters - indeed, they may be less likely, if members take the community discourse more seriously than so-called “trolls” and “flamers.” In addition, the obvious solution of disregarding anonymous posts might introduce more bias than it eliminates. Whatever statistical noise emanates from anonymous posting, presumably fuzzes equally both sides of the research hypothesis.

Coder Reliability

To introduce a measure of inter-coder reliability, a 200-comment discussion was analyzed by both coders and the results compared. Because the size of the “content universe” (Riffe et al, 1998) under study theoretically comprises the entire medium of Web site message-boards, the necessary volume of the reliability sample relative to the study population is incalculable. However, the reliability sample does amount to 10 percent of the study volume, allowing calculation of some standard reliability measures. The simple agreement figure between the two coders who participated in this content analysis is available in the data section below: they disagreed on about 5 percent of decisions in the control discussion.

A more complex method that attempts to correct for agreement by probability, a particularly important consideration with just two coders, is Scott’s Pi formula (Riffe et al, 1998). The formula is $Pi = \frac{\%OA - \%EA}{1 - \%EA}$, where OA = observed agreement and EA = expected agreement. Observed agreement on the control discussion was about 95 percent. To determine expected agreement, the total number of content items (posts) placed by the coders into each category (positive, neutral or negative) are converted into fractions.

Table 1: Value-response for control discussion

	Positive	Neutral	Negative	Total
Coder #1 (RMG)	7	54	89	158
Coder #2 (JBN)	8	58	92	158

Thus the coders collectively made 15 positive, 112 neutral and 181 negative judgments, of a total of 316. Thus each category comprises 4.7 percent, 35.4 percent and 57.3 percent of the total, respectively. Those percentages represent the probability that each particular agreement between coders was due to chance; to yield the expected agreement formula, each probability (expressed as a fraction of 1) is squared, and the sums calculated. The probabilities are .04, .354 and .573; when each is squared and summed, the total EA, or total agreement expected by chance alone, is .455, or 45.5 percent. Plugging in the simple observed agreement and the derived expected agreement into Scott's Pi formula yields the following:

$$Pi = \frac{.95 - .455}{1 - .455} = \frac{.495}{.545} = .908, \text{ or a Pi reliability of 90.8 percent. Thus the high OA of}$$

95 percent is "corrected" downward for the agreement due to chance alone. The relatively high EA value of 45.5 percent is a consequence of the three-increment axis used.

Analysis - Thread Depth

Table 2: Average thread depth for original parent posts

Token	Token Original Parents	Non-token Original Parents
Steam	2.6 posts	3.2 posts
iTunes	7 posts	4 posts
qwerty	2.8 posts	2.5 posts

Note that coder reliability figures above are not applicable to this step of the analysis; thread depth measurement is a purely mechanical process, yielding ratio-level data. Unfortunately, individual thread depth is not susceptible to meaningful analysis; it is only in the aggregate - within a single discussion - that the hypothesis predicts significant variance. This effectively limits the statistical tools that may be applied to thread depth: each discussion's measurement is a single frequency data point compared to the control population (non-token original parents).

A simple comparison of means in the table above, however, clearly shows equivocal results for the hypothesis that original parent posters employing the token attract more attention, represented by deeper threads. This straightforward observation, in fact, might be the clearest way of understanding thread depth measurement. For the largest sample group, the null hypothesis must be accepted; for the smaller samples it can only be rejected with great uncertainty. The degree to

which either hypothesis governs also is necessarily unclear.

Constructing a test with multiple data points by combining mean values from each discussion is easy, but a statistic such as a chi-square test runs into serious problems. Because the thread depths are mean values, they already incorporate standard variance, mitigating somewhat the existence of so few data points. More important, though, each thread depth average is effectively a separate population with separate expected frequency; conflating these mean values with real nominal data casts validity into question. Nevertheless, a standard chi-square may be computed for the mean values above: using $X^2 = \frac{(O_i - E_i)^2}{E_i}$, where O_i is observed frequency - the calculated mean - and E_i is expected frequency, or the control frequency. The chi-square formula yields

$$X^2 = \frac{(2.6 - 3.2)^2}{3.2} + \frac{(7 - 4)^2}{4} + \frac{(2.8 - 2.5)^2}{2.5} = 2.399. \text{ Assuming the standard desired}$$

probability of $\geq .05$, and a single degree of freedom (the means are nominal values), the chi-square significant difference threshold is 3.841 (Wimmer et al, 2000). Even setting aside the problems already outlined, the first part of null hypothesis I must be accepted; the prediction that token parents would engender deeper threads and, thus, more attention, is rejected.

Analysis - Support Response

Table 3: Results of support-response coding

Token	Steam	iTunes	qwerty
Token+	12 (25 percent)	15 (12 percent)	16 (14.7 percent)
Tokeno	22	54	46
Token-	14 (29.2 percent)	57 (45 percent)	47 (43 percent)
NT+	48 (7.1 percent)	28 (8 percent)	7 (6.2 percent)
NTo	369	186	55
NT-	263 (38.7 percent)	138 (39 percent)	51 (45 percent)

In every case, token posts more frequently drew positive response than the control group posts. However, in one of three cases token posts also drew proportionally more negative response; and token posts did not substantially reduce negative feedback. Because, as noted above, each sampled discussion constitutes a separate population, separate statistics determinations for each are helpful. A basic cross-tabulation can be applied to each set of sample data to determine whether token presence resulted in significant difference. For the first sample, “Steam,” the data matrix looks like this:

Table 4: Support-response matrix for token “Steam”

Steam	+	o	-
Token	12	22	14
Non-token	48	369	263

The token posts are expressed as a frequency within categories, e.g. $\frac{12}{60}$ for positive replies, etc.. Using the expected-frequency formula for cross-tab analysis, $E_{ij} = \frac{R_i C_j}{N}$, expected frequency is calculated for each cell. The results look like this, with each E_{ij} in the small right-hand boxes:

Table 5: Cross-tabulation of support response for token “Steam”

Steam	+		o		-	
Token	12	4	22	26	14	18
Non-token	48	56	369	365	263	259

The chi-square formula adapted to cross-tab analysis is $X^2 = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

(Wimmer et al, 2000). Applied to the Steam cross-tab matrix, the formula yields a chi-square of 18.75. Assuming probability of $\geq .05$, with two degrees of freedom, the Steam frequency distribution exceeds the threshold value of 5.991 - indeed,

exceeds the threshold value of 10.827 for $\geq .001$ probability. While the cross-tab analysis does not illuminate in which direction the distribution deviates, the simple percentage values above make clear that hypothesis II, that posts using the token would engender positive responses and, thus, greater credibility, is supported in the first sample discussion.

Similar analysis for the second and third samples follows.

Table 6: Support-response matrix for token “iTunes”

iTunes	+		o		-	
Token	15	11	54	63	57	x
Non-token	28	32	186	177	138	144

Applying the modified chi-square formula yields a value of 4.65, which falls short of significance given $\geq .05$ probability and two degrees of freedom. Token posts actually drew more negative responses proportionally than did non-token posts, and only slightly more positive responses. Research hypothesis II rejected.

Table 7: Support-response matrix for token “qwerty”

qwerty	+		o		-	
Token	16	11	46	50	47	48
Non-token	7	12	55	51	51	50

Applying the modified chi-square formula yields 5.03, short of the 5.991 threshold assuming $\geq .05$ probability and two degrees of freedom. Research hypothesis II rejected.

Discussion

As briefly noted above, the level of noise present in the discussion samples seemed exceptionally high. However dedicated or involved online community members may be, they clearly suffer from an inundation of meaningless chatter interspersed with their messages. The noise level encountered in these discussions, and in others which were not amenable to analysis, proved more insidious and obstructive than its passive static or lost signal namesakes; noise online is active, it seeks attention and drowns out meaningful messages. As numerous as were textless posts and irrelevant scribbles, more so were messages targeted at inciting anger or indignation; the infamous “flames” and “trolls” of cyberspace.

In some instances, in fact, group members employed tokens to counter misunderstanding and confusion set off by flammers and trolls. In others, flammers and trolls sometimes used tokens, as well as other strategies like false facts and misleading arguments, to cloud the discussions studied, resulting in even more noise from genuinely confused members or newcomers. Even group members in the communities studied were seen to wonder aloud, why do flammers devote so much time to their craft? Posting tired, familiar flames with racist or homophobic or evangelist themes (all of which are common) takes little time or imagination. But some flammers went to unusual lengths simply to throw spanners into the

discussion works; one example is the “paste-and-change” flamer, who ostensibly posts and attributes outside source material (e.g. an article under discussion or related material) as a service to other readers, while subtly changing the posted text to reverse or distort its meaning. Another type of noise is the wide-eyed flamer who poses an apparently innocent question that is well-known to the community to be “flamebait,” such as “Is Microsoft dominance really such a bad thing?” or “I don’t see anything wrong with downloading music for free,” two examples from the analyzed discussions.

Largely, though, the communities studied confidently ignored noise-makers and engaged each other. The built-in user moderation systems, which most members surely employ to minimize their reading workload, played a role in emphasizing meaningful information and pushing garbage to the bottom of the pile. In an environment of constant distraction, communicators must necessarily employ powerful strategies for gaining and maintaining attention and trust. Those moderation systems, however, powerfully skew the experiences of community members by substituting the judgment of moderators - generally a subset of members, as on Slashdot and Kuro5hin, or site administrators for other groups - for the members’ own on the majority of submissions. An online community is much more like an echo chamber than an empty theater, and user moderation seems likely to produce a strong conservative inertia against external ideas and

individuals and in favor of established ones.

Obviously, the interplay of members among themselves and the consequences for credibility interpretation through positive/negative evaluation promise extreme complexity. Again, however, this is the way of dynamic systems, to arrange themselves to maximize utility for their users, and such complexity is nothing new in RL. The meetings of a large social club or bureaucratic organization might function in a similar manner, with most input routed through gatekeeper members, and the resultant stream of information would both reflect and strengthen the judgments of membership as to particular speakers' credibility and authority. No one would claim that RL group dynamics are simply an intractable problem, because entire fields like group and organizational psychology, visual/verbal communication and rhetoric are long used to the alphabet of RL group interaction. Understanding tokens, which function as the "missing variables" of CMC group dynamics, requires similar engagement with online textual manipulation as more than merely signposts in cyberspace, and an attempt to dissect online group dynamics as an area accessible to quantitative research, not the perception of CMC as a kind of black art that has driven nearly all work in the last five years into less-rigorous, qualitative methods.

A recent attempt to explore the origins of flaming offers an insightful explanation that points in the direction of understanding CMC textual cues as

instrumental in nature. After criticizing the limitations model of CMC for the well-known drawbacks outlined above, Thompsen (2003) offers a “social influence model of flaming” that recognizes the importance of group interaction in message interpretation. Understanding the cues messages carry prevails over context-free “objective” observation in Thompsen’s schema. Summing up his argument for context-dominated understanding of flaming, Thompsen wonders if “a flame is not a flame until someone calls it a flame” (334). Contextual understanding of flaming accords with a model of textual cues - tokens - as manipulative methods of evoking particular responses, methods that likely depend on the community under study. Thompsen concludes by calling on CMC workers to leave limitations-model preconceptions behind and understand textual manipulation online as interactional, rather than mechanical: “... [R]esearchers should move toward examining flaming not simply as an effect of communication technology, but as a reflection of the social negotiation of meaning through communication media” (343).

An enticing discussion of precisely the kind of low-level tokens studied in this analysis is Lee’s (2003) exploration of “codes in cyberspace,” a discussion of jargon and “slanguage” used on e-mail. Lee’s observations took place on a tech-centric e-mail list (*Technoculture*), and it’s telling that she notes the commonality of acronyms like MUD (Multi-User Dungeon), MOO (Multi-user Object-Oriented) and MUSE (Multi-User Simulated Environment); tokens (Lee does not use the

word) very rarely found outside jargon-heavy circles. Tantalizingly, Lee writes that “... [J]argon helps create and define e-mail communities, and new members embrace the lingo as signs of belonging” (319).

Lee also makes an interesting comparison between affectations in text - in one case, affecting a drawling Western accent with spellings like “Wahl, I kin see we don’t want no FED’RAL AGENTS on this list” - and personal performance, visual or otherwise, before a group: “E-Mail as Performance Space” (322). This idea of the CMC message as conscious presentment, to a particular audience, of a particular identity (cf. “the message as mask,” see Danet, 1998; and Wiszniewski and Coyne, 2002) *via textual manipulation* virtually implies the question of what particular manipulations, or tokens, are employed to what rhetorical ends.

While the analogy of CMC text as performance or dance is an elegant, appealing one, it is not enough. It may seem a blindingly obvious, perhaps meaningless question, to ask why CMC group members say what they say, but here it is the core question: why do they dance? We as humans and as communications workers ask that kind of question instinctively in RL: why does a speaker gesture as he does? Why do fast-food restaurants paint their franchises in bright primary colors? Why does Louis-Ferdinand Céline end sentences with ellipses, or William Faulkner sometimes not end them at all? We are in the infancy of our understanding of tokens, the functional equivalents of style, inflection,

gesture in CMC, and asking “why?” of a chat-room line, a message-board post or an e-mail may feel foreign, alien to our experience. The very dichotomy between CMC communications and “RL,” as if what we say online somehow does not exist, points to a troubling undervaluation of electronic communication, especially in this age of its growing dominance.

On a related note, while the operation of tokens as credibility strategies is left an open question by this analysis’ conclusions, it must be noted that the range of factors which cannot be accounted for in this type of research is daunting. Of course posters employ traditional rhetorical strategies, as communicators do in all other media, along with some that may prove unique to CMC: careful attention to spelling and grammar, on a basic level, and principles of logic and rhetorical organization on a higher one. The identity of posters, when it can be known, is an open question so far as credibility is concerned. Reliably tracking who knows whom, and who feels predisposed to trust whom, and who not, within a single community alone is nearly as intimidating a prospect as studying the entire corpus of online communities.

The vastly diverse character of communities themselves, which may rival their RL counterparts in breadth if not in depth, certainly plays a role in which credibility strategies are given more weight to counter the greater ease of communicating online. Speaking to a large CMC group does not require

overcoming many of the barrier factors that oppose us in real life, like the cost of advertising or programming and media gatekeeping.

In addition, CMC lexicons pose a truly frightening problem which goes hand in hand with their uniquely powerful role within their specific communities. “For something to be a signifier, it has to be located within a code in which its uniqueness, its difference from any other possible signifier, can be recognized” (Grossberg et al., 1998, p. 132). By their nature CMC groups develop internal and exclusive lexicons - libraries of tokens - which are more effective the more they diverge from the vocabulary of other groups. The net effect is not unlike a family of languages that actively develop in divergent directions, further complicating the researcher’s task. Isolating meaningful tokens, after all, presupposes an understanding of what group members are saying. Just as we presume understanding of the nuance of dance in a theater critic, we presume in ourselves at least cursory knowledge of manipulative and rhetorical strategies in language.

This facility allows us to skirt rigidity and ascribe adjectives like “friendly” or “supportive” or “harsh” or “condemnatory” to RL or CMC communications without recourse to semiotics. Presuming that we can, without further effort, do the same in the various coded languages of cyberspace, whose purpose is exclusion of outsiders (including communications researchers), may be seriously misplaced confidence. Tacit recognition of the sheer incomprehensibility or apparent

intractability of CMC codes may also explain the sparse literature directly addressing tokens like obscure acronyms, emoticons and jargon.

Regardless of these factors, I maintain tokens do play the instrumental role pointed to by early research in this field, and I am confident that further investigation will uncover the means of using them as information filters and credibility signifiers. The equivocal results of this content analysis, I feel, are a product of confounding variables, small sample sizes, insufficient communities and discussions studied, and, especially, inability to judge the relative size of token effects compared to traditional rhetorical strategies. These criticisms are not meant to reflect a false modesty, but a genuine frustration at the sheer magnitude of the difficulty of investigating token activity even within single communities.

Part of that problem is logistics; more extensive and more rigorous analysis undoubtedly will give greater confidence in statistical results. But I am not sure those results would differ much from those reached here. The bigger problem is the simple absence of prior development in this field, and the absence of direct comparison between traditional rhetoric and rhetoric via textual manipulation. For years CMC workers have known that online communities, like RL ones, develop particular manners of speaking and unique terminology. In traditional, FtF communication these developments are called slang, patois, jargon and “in-group” speech, and they’re extensively studied both as indicators of group development

and according to their use by particular speakers for particular purposes, whether attracting attention, proclaiming member status and credibility, or exclusion of outsiders. Why, then, has CMC not recognized virtually the same phenomenon, and explored why exactly people say what they say online?

While this investigation has proceeded strictly along content-analytical lines, it must be acknowledged that the foundational idea involved - the existence and instrumental activity of tokens - emerges from an holistic and admittedly anecdotal investigation. The choice of tokens for these analyses emerged from the familiarity of the author with the communities and the subjects involved, and the choices bear the imperfections of his understanding. An encompassing token typology, as noted above, is unavailable, and indications are that development of one will face very significant hurdles until a better understanding of tokens, and of CMC generally, is attained. More research like this will certainly help - but the theoretical background for instrumental tokens is simply missing. What have been dismissed or marginalized as merely signifiers of community presence are in this analysis elevated to rhetorical tool.

Why These Tokens?

All three tokens chosen were single words; two of them - iTunes and Steam - are recognizable brands (to their respective communities), and the other - qwerty - is an acronym. As discussed above, token use seems to be most prevalent in

techno-centric communities, and each of the studied tokens concerns people who work and play in techno-centric fields. Like any community signifier, tokens should be used by communities encompassing a broad range of sizes, interests and occupations; the tokens studied correspond to communities of greatly varying composition.

iTunes, the Apple Computer-branded online music store and music jukebox software, is probably familiar to anyone who downloads music legally from the Internet; the iTunes Music Store is that market's leader. Thus the potential community size to look for use of iTunes as a token is very large. Conversely, the value of the word as a token may be diminished by the broad base of people who have heard it and recognize it. At the opposite end of that size spectrum, only the limited group of people who play online video games by Valve Software, Inc., are in a position to recognize 'Steam' in the context of online gaming. Because that community is small and insular, the experimental value of using 'Steam' as a study token is correspondingly high. Unsurprisingly, 'Steam' exhibited highly significant correlation between token use and credibility. More details on the nature of each token are given in the appendix.

Recognizing the limitations of quantitative validity, I have nevertheless chosen to study specific tokens at the lowest level they may operate, a single fleeting burst of communication among community members. Initially choosing

tokens to study, then, resembles case study analysis - on an informal level, to be sure - more than rigorous definition of testable data. The tokens “Steam,” “qwerty” and “iTunes” were chosen from plain observation of what group members appeared to be saying to each other, and what terms they included in their arguments to appear credible. Determining which tokens to test subjectively was necessarily grounded in the author’s experience with the communities under study, and as noted above, the tokens themselves might be the largest variable involved in the analysis.

Despite confounding factors, rejection of the credibility hypothesis in two of three discussions was marginal; and significance met and exceeded acceptable levels in the first sample. Arriving at such intriguing results in the face of simply massive theoretical hurdles actually exceeds the author’s expectations, and clearly indicates the need to investigate further. All things considered, this analysis demonstrates that tokens and credibility have some relationship in CMC. A review of the literature shows that result is novel to the field, and it poses all kinds of interesting questions that only more in-depth research can answer. This author believes the field is even now struggling toward exactly that kind of engagement with tokens as rhetorical device.

Stake (1995) tried to outline the balance between approaching research targets from outside hard empirical underpinnings:

To sharpen the search for explanation, quantitative researchers perceive what is happening in terms of descriptive variables.... To sharpen the search for understanding, qualitative researchers perceive what is happening in key episodes or testimonies, represent happenings with their own direct interpretation and stories (i.e., narratives) (40).

Because the concept of tokens is so clearly in action within online communities, yet the mechanics of their work is as variable and obscure as the Internet itself, it is exceedingly difficult either to particularize a single discussion or generalize to a larger community or the online world as a whole. The topic of instrumental tokens is so indistinct, perhaps, because of their ubiquity in CMC and the unprecedented rapidity with which they arise and develop. Perhaps workers in the field, though they have grown well beyond the circumscribed limitations model, still have not learned to examine online communication with the same depth of field they apply to words that issue from our mouths into the air, or our pens onto paper.

This theoretical review and brief, equivocal analysis certainly will not serve to define tokens, to pin them down or explain their origins or function. But it may (and I have no illusions there will be a cause and effect relationship) be an early effort into a field that will only grow in importance as time passes and CMC exerts further influence on the development of our language and its use.

References

- Anderson, B. (1983). *Imagined communities: reflections on the origin and spread of nationalism*. London: Verso.
- Avery, C. & Zeckhauser, R. (1997). Recommender systems for evaluating computer messages. *Communications of the ACM*, 40, 88-.
- Baym, N. (1993). Interpreting soap operas and creating community: Inside a computer-mediated fan culture. *Journal of Folklore Research*, 30, 144-176.
- Baym, N. (1995). The emergence of community in computer-mediated communication. In Jones, S.G. (Ed.), *Cybersociety: Computer-mediated communication and community* (138-163). Thousand Oaks, CA: Sage.
- Baym, N. (1996). Agreements and disagreements in a computer-mediated discussion. *Research on Language and Social Interaction*, 29, 315-345.
- Baym, N. (1998). The emergence of on-line community. In Jones, S.G. (Ed.), *Cybersociety 2.0: Revisiting computer-mediated communication and community* (35-68). Thousand Oaks, CA: Sage.
- Bennett, T. (1982). Media, 'reality', signification. In Gurevitch, M., Bennett, T., Curran, J. & Woollacott, J. (Eds.), *Culture, society and the media* (287-308).
- Bolter, J.D. (2001). *Writing Space: Computers, hypertext and the remediation of print*. Mahwah, NJ: Lawrence Erlbaum.
- Bolter, J.D. (2003). Virtual reality and the redefinition of self. In Strate, L., Jacobson, R.L. & Gibson, S. (Eds.), *Communication and Cyberspace: Social interactions in an electronic environment* (123-137). Cresskill, NJ: Hampton.
- Collot, M. & Belmore, N. (1996). Electronic Language: A new variety of English. In Herring, S.C. (Ed.), *Computer-Mediated Communication: Linguistic, social and cross-cultural perspectives* (13-28). Amsterdam: John Benjamins.
- Connery, B. (1997). IMHO: Authority and egalitarian rhetoric in the virtual coffeehouse. In Porter, D. (Ed.), *Internet Culture* (161-179). New York: Routledge.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Curran, J., Gurevitch, M. & Woollacott, J. (1982). The study of the media: Theoretical approaches. In Gurevitch, M., Bennett, T., Curran, J. & Woollacott, J. (Eds.), *Culture, society and the media* (11-29).

- Danet, B. (1998). Text as mask: Gender, play, and performance on the internet. In Jones, S.G. (Ed.), *Cybersociety 2.0: Revisiting computer-mediated communication and community* (129-158). Thousand Oaks, CA: Sage.
- Davidson, A.L. & Schofield, J.W. (2002). Female voices in virtual reality: Drawing young girls into an online world. In Renninger, K.A. & Shumar, W. (Eds.), *Building Virtual Communities: Learning and change in cyberspace* (34-59). Cambridge: Cambridge University Press.
- Davis, B.H. & Brewer, J.P. (1997). *Electronic Discourse: Linguistic individuals in virtual space*. Albany, NY: State University of New York Press.
- Elgesem, D. (1996). Privacy, respect for persons, and risk. In Ess, C. (Ed.), *Philosophical Perspectives on Computer-Mediated Communication* (45-66). Albany, NY: State University of New York Press.
- Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication*, 8, 8-34.
- Foster, D. (1997). Community and identity in the electronic village. In Porter, D. (Ed.), *Internet Culture* (23-37). New York: Routledge.
- Galegher, J., Sproull, L. & Kiesler, S. (1998). Legitimacy, authority, and community in electronic support groups. *Written Communication*, 15, 493-.
- Geertz, C. (1973). *The Interpretation of Cultures*. New York: Basic.
- Giese, M. (1998). Constructing a virtual geography: Narratives of space in a text-based environment. *Journal of Communication Inquiry*, 22, 152-.
- Grossberg, L., Wartella, E. & Whitney, D.C. (1998). *MediaMaking: Mass media in a popular culture*. Thousand Oaks: Sage.
- Gurak, L.J. (1997). *Persuasion and Privacy in Cyberspace: The online protests over Lotus MarketPlace and the Clipper Chip*. New Haven: Yale University Press.
- Haythornthwaite, C. (2002). Building social networks via computer networks: Creating and sustaining distributed learning communities. In Renninger, K.A. & Shumar, W. (Eds.), *Building Virtual Communities: Learning and change in cyberspace* (159-190). Cambridge: Cambridge University Press.
- Heim, M. (1993). *The Metaphysics of Virtual Reality*. New York: Oxford University Press.
- Herring, S. C. (1996). Posting in a different voice: Gender and ethics in computer-mediated communication. In Ess, C. (Ed.), *Philosophical Perspectives on Computer-Mediated Communication* (125-145). Albany, NY: State University of New York Press.

- Hoadley, C. & Pea, R.D. (2002). Finding the ties that bind: Tools in support of a knowledge-building community. In Renninger, K.A. & Shumar, W. (Eds.), *Building Virtual Communities: Learning and change in cyberspace* (321-354). Cambridge: Cambridge University Press.
- Hudson, R.A. (1980). *Sociolinguistics*. Cambridge: Cambridge University Press.
- Kiesler, S., Siegel, J. & McGuire, T.W. (1984). Social psychological aspect of computer-mediated communication. *American Psychologist*, 39, 1123-1134.
- Kolko, B. & Reid, E. (1998). Dissolution and fragmentation: Problems in on-line communities. In Jones, S.G. (Ed.), *Cybersociety 2.0: Revisiting computer-mediated communication and community* (212-229). Thousand Oaks, CA: Sage.
- Krout, M.H. (1971). The symbolism of objects and movements. In Bosmajian, H. (Ed.), *The Rhetoric of Nonverbal Communication*. Glenview, IL: Scott, Foresman.
- Lea M. & Spears, R. (1991). Computer-mediated communication, de-individuation and group decision-making. *International Journal of Man-Machine Studies*, 34, 283-301.
- Lee, J.Y. (2003). Charting the codes of cyberspace: A rhetoric of electronic mail. In Strate, L., Jacobson, R.L. & Gibson, S. (Eds.), *Communication and Cyberspace: Social interactions in an electronic environment* (307-328). Cresskill, N.J.: Hampton.
- Levinson, R.W. (1988). *Information and Referral Networks: Doorways to human services*. New York: Springer.
- Liu, Y. (1997). Authority, presumption, and invention. *Philosophy and Rhetoric*, 30, 413-442.
- Lyon, D. (1997). Cyberspace sociality: Controversies over computer-mediated relationships. In Loader, B.D. (Ed.), *The Governance of Cyberspace* (23-37). New York: Routledge.
- McLuhan, M. (1964). *Understanding Media: The extensions of man*. New York: McGraw-Hill.
- McKenna, K.Y.A., Green, A.S. & Gleason, M.E.J. (2002). Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues*, 58, 9-.
- Meyrowitz, J. (1985). *No Sense of Place: The impact of electronic media on social behavior*. New York: Oxford University Press.
- Miller, S.E. (1996). *Civilizing Cyberspace: Policy, power, and the information superhighway*. New York: Association for Computing Machinery Press.
- Mitra, A. (1997). Virtual community: Looking for India on the Internet. In Jones, S.G. (Ed.), *Virtual culture: Identity and communication in cybersociety* (55-79). Thousand Oaks,

CA: Sage.

Murray, D.E. (1991). The composing process for computer conversation. *Written Communication*, 8, 35-55.

O'Keefe, D.J. (1990). *Persuasion: Theory and Research*. Newbury Park, CA: Sage.

Parks, M.R. & Floyd, K. (1996). Making friends in cyberspace. *Journal of Communication*, 46, 80-.

Raymond, E.S. (2000). *The Jargon File*. Available at <<http://catb.org/esr/jargon/>>.

Rheingold, H. (1993). *The Virtual Community: Homesteading on the electronic frontier*. Reading, MA: Addison-Wesley.

Rheingold, H. (1998). Virtual communities. In Hesselbein, F., Goldsmith, M., Beckhard, R. & Schubert, R.F. (Eds.), *The community of the future* (115-122). San Francisco: Jossey-Bass.

Riffe, D., Lacy, S. & Fico, F. (1998). *Analyzing Media Messages: Using quantitative content analysis in research*. Mahwah, N.J.: Lawrence Erlbaum.

Shank, G. & Cunningham, D. (1996). Mediated phosphor dots: Toward a post-Cartesian model of computer-mediated communication via the semiotic superhighway. In Ess, C. (Ed.), *Philosophical Perspectives on Computer-Mediated Communication* (27-41). Albany, NY: State University of New York Press.

Shaw, D.F. (1997). Gay men and computer communication: A discourse of sex and identity in cyberspace. In Jones, S.G. (Ed.), *Virtual culture: Identity and communication in cybersociety* (133-145). Thousand Oaks, CA: Sage.

Sproull, L. & Kiesler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.

Stake, R. (1995). *The Art of Case Study Research*. Thousand Oaks: Sage.

Thompson, P.A. (2003). What's fueling the flames in cyberspace? In Strate, L., Jacobson, R.L. & Gibson, S. (Eds.), *Communication and Cyberspace: Social interactions in an electronic environment* (329-347). Cresskill, N.J.: Hampton.

Wilbur, S. (1997). An archaeology of cyberspaces: Virtuality, community, identity. In Porter, D. (Ed.), *Internet Culture* (5-22). New York: Routledge.

Wilkins, H. (1991). Computer talk: Long-distance conversations by computer. *Written Communication*, 8, 56-78.

Wimmer, R. & Dominick, J. (2000) *Mass Media Research: An Introduction*. Belmont, CA: Wadsworth.

Wiszniewski, D. & Coyne, R. (2002). Mask and identity: The hermeneutics of self-construction in the information age. In Renninger, K.A. & Shumar, W. (Eds.), *Building Virtual Communities: Learning and change in cyberspace* (191-214). Cambridge: Cambridge University Press.

Yates, S.Y. (1996). Oral and written linguistic aspects of computer conferencing: A corpus-based study. In Herring, S.C. (Ed.), *Computer-Mediated Communication: Linguistic, social and cross-cultural perspectives* (29-46). Amsterdam: John Benjamins.

Appendix: Raw Data with Source Notes

Control Discussion

“Links to Tens of Thousands of Legal Music Downloads”

Available online at:

<<http://www.kuro5hin.org/story/2003/9/5/05113/70314>>

Discussion volume: 197 comments.

Token under examination: “riaa”

Synopsis: Kuro5hin.org posted this discussion on Friday, Sept. 5.

This discussion followed a long essay about music downloading, fair use and alternative channels of music distribution. Gravitating towards the implications of the music-download business model for fair use rights and privacy online, many posters hurled accusations at the RIAA, or Recording Industry Association of America, regarding that group’s public relations and government lobbying efforts.

Results:

Thread Depth

Both coders registered the same number of original parent posts (39), and divided them among token-using (7) and non-token-using (32) identically. This coding step was objective.

Support Response

Both coders registered the same number of token-post replies (16) and non-token-post replies (142). This coding step was objective.

Of 158 decisions regarding response-value coding, coders disagreed on 8, or about 5 percent. This coding step represented judgment along the three-increment scale.

Sample Discussion #1:

“Half Life 2 Source Code Leaked”

Available online at:

<<http://games.slashdot.org/article.pl?sid=03/10/02/1547218&mode=thread&tid=126&tid=127&tid=156&tid=186&threshold=-1>>

Discussion volume: 1,027 comments.

Token under examination: “steam”

Synopsis: Slashdot posted this discussion on Thursday, Oct. 2.

“*Half Life 2*” is the upcoming, unreleased sequel to a popular video game; this story covered news that the under-development source code for the game had been stolen and leaked onto the Internet. According to the developer’s Web site, “Steam” is the online multiplayer service used for the original game and projected for use in the sequel; the name is a play on that of the developer, Valve Software. Much of the community’s concern over this leak revolved around use of this leaked code to compromise security features of the Steam service and render cheating and game exploits easier.

Posters made arguments about the code leak’s potential effects on the Steam service from a variety of perspectives, but analysis bore out the anecdotal conclusion that observations on the leak’s ramifications for Steam were critiqued with some deference by the community.

Results:

Thread Depth

19	Parent posts containing token
249	Parent posts not containing token
2.6	Average thread depth of token parents
3.2	Average thread depth of non-token parents

Support Response

680	Replies to non-token posts
48	Replies to token posts
48	Positive replies to non-token posts
369	Neutral replies to non-token posts
263	Negative replies to non-token posts

Of replies to non-token posts, 7.1 percent were positive.
Of replies to non-token posts, 38.7 percent were negative.

12	Positive replies to token posts
22	Neutral replies to token posts
14	Negative replies to token posts

Of replies to token posts, 25 percent were positive.
Of replies to token posts, 29.2 percent were negative.

Coder error: Total number of posts analyzed falls short of the Web site's reported count by 31 (996 analyzed of 1,027 reported), or about 3 percent.

Sample Discussion #2:

“New Napster Off To A Solid Start”

Available online at:

<<http://slashdot.org/article.pl?sid=03/11/03/%201436236&mode=thread&tid=141&tid=188&tid=98&tid=99&threshold=-1>>

Discussion volume: 593 comments.

Token under examination: “iTunes”

Synopsis: Slashdot posted this essay and invited discussion on Monday, Nov. 3.

This discussion followed news of a re-branded Napster music download service and ensuing debate about various music download services, the economics of online content industries and the morality of illegally copying music online. While the discussion ostensibly was aimed at the Napster service, a number of

posters quickly began comparing Napster and the download market leader, Apple Computer's iTunes Music Store.

While the tenor of the discussion as a whole tended highly negative, posters addressing iTunes' strengths and weaknesses garnered slightly more praise; they also attracted more criticism than other posters.

Results:

Thread Depth

31	Parent posts containing token
72	Parent posts not containing token
7	Average thread depth of token parents
4	Average thread depth of non-token parents

Support Response

352	Replies to non-token posts
126	Replies to token posts
28	Positive replies to non-token posts
186	Neutral replies to non-token posts
138	Negative replies to non-token posts

Of replies to non-token posts, 8 percent were positive.

Of replies to non-token posts, 39 percent were negative.

15	Positive replies to token posts
54	Neutral replies to token posts
57	Negative replies to token posts

Of replies to token posts, 12 percent were positive.

Of replies to token posts, 45 percent were negative.

Coder error: Total number of posts analyzed falls short of the Web site's reported count by 12 (581 analyzed of 593 reported), or about 2 percent.

Sample Discussion #3:

“An Argument for Dvorak”

Available online at:

<<http://www.kuro5hin.org/story/2003/8/28/21547/2069>>

Discussion volume: 302 comments.

Token under examination: “qwerty”

Synopsis: Kuro5hin.org posted this discussion on Friday, Aug. 29.

This discussion revolved around a short essay, submitted by a Kuro5hin.org contributor, advocating an alternative key-mapping arrangement for computer keyboards, the “Dvorak” keyboard, for speedier typing. The Dvorak key-mapping is an alternative to the traditional “qwerty” keyboard inherited from typewriters.

In a discussion on worker ergonomics or office efficiencies, either “Dvorak” or “qwerty” would have made good candidates for analysis. In this case, posters who appeared authoritative debated the merits of the Dvorak and qwerty key-mappings directly; others were more likely to give personal accounts of stress injuries or question the need for a debate on replacing the standard keyboard mapping.

Results:

Thread Depth

39	Parent posts containing token
30	Parent posts not containing token
2.8	Average thread depth of token parents
2.5	Average thread depth of non-token parents

Support Response

113	Replies to non-token posts
109	Replies to token posts

7	Positive replies to non-token posts
55	Neutral replies to non-token posts
51	Negative replies to non-token posts

Of replies to non-token posts, 6.2 percent were positive.
Of replies to non-token posts, 45 percent were negative.

16	Positive replies to token posts
46	Neutral replies to token posts
47	Negative replies to token posts

Of replies to token posts, 14.7 percent were positive.
Of replies to token posts, 43 percent were negative.

Coder error: Total number of posts analyzed falls short of the Web site's reported count by 11 (291 analyzed of 302 reported), or about 4 percent.

Vita

A resident of Baton Rouge, Louisiana, Ryan Goudelocke attended the Episcopal School of Acadiana in Cade, Louisiana, until 1997. He graduated *magna cum laude* in three years from Oglethorpe University, Atlanta, Georgia, with a Bachelor of Arts degree in politics. After eighteen months of journalism graduate study, he joined The Advocate newspaper in Baton Rouge on the crime beat. He left the newspaper in May 2003 to complete his master's degree in mass communication and to enter Louisiana State University's Paul M. Hebert Law Center, where he is currently a second-year student.