MIS Quarterly

SPECIAL ISSUE: IS FOR SYMBOLIC ACTION

TAKE THEIR WORD FOR IT: THE SYMBOLIC ROLE OF LINGUISTIC STYLE MATCHES IN USER COMMUNITIES[1]

Stephan Ludwig
Department of Management, University of Westminster, 309 Regent Street,
London W1B 2HW, UNITED KINGDOM {s.ludwig@westminster.ac.uk}

Ko de Ruyter, Dominik Mahr, Martin Wetzels, and Elisabeth Brüggen
Department of Marketing and Supply Chain Management, Maastricht University,
P.O. Box 616, 6200, MD Maastricht, THE NETHERLANDS
{k.deruyter@maastrichtuniversity.nl} {d.mahr@maastrichtuniversity.nl}
{m.wetzels@maastrichtuniversity.nl} {e.bruggen@maastrichtuniversity.nl}

Tom de Ruyck
Head of Consumer Consulting Boards at InSites Consulting, Evergemsesteenweg 195,
9032 Ghent, BELGIUM {tom.deruyck@insites-consulting.com}

User communities are increasingly becoming an essential element of companies' business processes. However, reaping the benefits of such social systems does not always prove effective, often because companies fail to stimulate members' collaboration continuously or neglect their social integration. Following communication accommodation theory, the authors posit that members' communication style alignment symbolically reflects their community identification and affects subsequent participation behavior. This research uses text mining to extract the linguistic style properties of 74,246 members' posts across 37 user communities. Two mixed multilevel Poisson regression models show that when members' linguistic style matches with the conventional community style, it signals their community identification and affects their participation quantity and quality. Drawing on an expanded view of organizational identification, the authors consider dynamics in members' social identification by examining trends and reversals in linguistic style match developments. Whereas a stronger trend of alignment leads to greater participation quantity and quality, frequent reversals suggest lower participation quantity. At a community level, greater synchronicity in the linguistic style across all community members fosters individual members' participation behavior.

Keywords: Linguistic style match (LSM), user communities, text mining, organizational identification, argument development quality

Introduction

An increasing number of companies assume that user communities can be leveraged to provide access to their end users' insights and resources (Dahlander and Frederiksen 2012). However, the benefits of this assumption have not always been realized; at least half of newly established user communities fail to sustain the quantity and quality of their members' participation (Ransbotham and Kane 2011). With virtually no barriers to switching to various alternative user communities (Iriberri and Leroy 2009), members' social integration and identification with the community is the key reason they stay and participate (Kohler et al. 2011).

[1]Mark Aakhus, Pär J. Ågerfalk, Kalle Lyytinen, and Dov Te'eni were the accepting senior editors for this paper.

Yet the socialization in these communities remains largely under the radar due to shortcomings of current approaches that lack the ability to adequately capture members' beliefs and perceptions based on surface-level parameters, such as number of posts and visit counts. Similarly, surveys in these settings suffer from low response rates and hence cannot provide exhaustive access to members' beliefs and perceptions. These approaches fail to recognize that socialization depends largely on communicative processes, such as language use (Moran and Gossieaux 2010). Therefore, it is important to take into account sociolinguistic factors that are embedded in language use and recognize these as a valuable means for examining how members integrate in communities and how communities evolve through interaction.

New managerial insights on members' social integration and participation in user communities may, however, emerge from their communication styles in their posts (Herring 2001). While users contribute their ideas and views via such posts, the styles in which they are formulated also serve as *in vivo* tests, providing information about consensual behavior and the extent to which members agree to it (Fayard and DeSanctis 2010). Although user communities thrive on diversity in members' ideas and views (Di Gangi et al. 2010), patterns of subtle similarities in the way members converse may thus indicate an active process of social integration, which drives members' ongoing participation.

Accordingly, this study incorporates emerging, cross-disciplinary theorizing that views communication as a symbolic action (Heracleous and Marshak 2004). The basic premise of this paradigm is that the act of communication transcends information content by attaching conversants' subjective meaning to situations (Heracleous and Marshak 2004) and conveying their underlying intentions (Fayard and DeSanctis 2010). In the body of research referred to as communication accommodation theory (CAT) (Giles 2009), a focal communicative process pertains to how conversants align with a consensual style of interaction. Accumulating evidence indicates that in computer-mediated, text-based communication, linguistic style matches (LSMs) among conversants signal social identification perceptions (Ireland and Pennebaker 2010) and influence cooperative behavior (Huffaker et al. 2011). However, this evidence and recent theorizing in CAT is primarily limited to communication dyads and cross-sectional designs. It remains uncertain whether LSM in a user community context reflect a process of social identification (Ireland and Pennebaker 2010). Furthermore, it is not clear how LSM, as a process, develops over time and whether diverging temporal patterns in LSM distinctly foster community members' participation behavior. Finally, communication research thus far has focused on text or a member posting text in one community but neglected the societal context in which the text is nested (Hardy et al. 2005). Thus, it is uncertain how differences between communities with regard to linguistic style practices influence the individual members' allocation of their efforts (Brockman and Morgan 2006). This study contributes to an emerging but limited body of research on the symbolic role of communication styles by addressing three critical issues.

First, it extends CAT by conceptualizing LSM in text-based user communities as a symbolic action that signals member identification with the community and predicts their participation efforts in the community. In contrast to the conventional focus on content words, LSM builds on function words which comprise pronouns, prepositions, articles, conjunctions, or auxiliary verbs; are used subconsciously; and constitute 55 percent of daily word usage.

We argue that these function words serve as a subtle, subconscious way to construct shared meaning among community members. We disentangle LSM's effect on two key community member performance parameters: (1) participation quantity (post frequency) and (2) participation quality (substantiated development of arguments). Both are critical for sustaining the value and appeal of the user community to its members and hosts (Ransbotham and Kane 2011). Second, in accordance with the inherently dynamic process of members' social integration into organizations (Levine and Moreland 1994), this study considers the temporal development, or gestalt, of LSMs. We suggest two distinct temporal patterns in members' LSMs (trend and reversals) and detail their distinct influence on members' participation efforts. Third, for the plethora of user communities, which suffer from a relative paucity of contextual implications for communication, it seems both managerially and conceptually relevant to examine between-community differences in members' LSM cohesiveness and their effects on individual members' contribution behavior.

Conceptual Background

With the rise of social media, user communities have become a more widely embraced communication format (Ransbotham and Kane 2011). Members' ongoing participation in terms of quantity and quality, crucial for the viability of user communities, can only be ascertained through the successful socialization of members into the community (Wasko and Faraj 2005). Participation quantity refers to the number of a member's posts, where a higher number increases the likelihood that the member raises topics that attract a larger audience and reciprocal participation (Wasko and Faraj 2005). Conversely, lack of sufficient participation quantity may make a community unappealing for members and lead online collectives to become more like ghost towns, deserted of virtual life (Ransbotham and Kane 2011). Although important, participation quantity still may be insufficient for sustaining the community if posts are merely undeveloped statements and opinions (Hansen and Haas 2001). In this vein, participation quality refers to members' efforts to develop their communication, which leads to better group discussion outcomes (Gouran 1990). This conceptualization corresponds with communicative argumentation quality research, which demonstrates that the development of reasoning, rather than its strength, predicts decision outcomes (Seibold et al. 2010). Thus, a community member contributing posts with more developed arguments achieves higher participation quality, which enhances the appeal of participation for others (Seibold and Meyers 2007). Notably, this measure of argumentation development allows comparisons between members' participation quality across communities, irrespective of the content.

In the absence of face-to-face contacts and other explicit member information, the words used in text-based posts become the only symbols available to assess members' socialization development (Kohler et al. 2011; Ma and Agarwal 2007). Longitudinal surveys are simply impractical in such settings and tracking quantitative surrogates, such as post counts, does not suffice. New insights may derive from research on communication accommodation, where people's degree of accommodation in linguistic style may be particularly relevant in computer-mediated communication settings, to assess conversants' social identification and subsequent behaviors. The central premise of communication accommodation theory (CAT) is that symbolic actions, inherent to communication styles, form the basis for linguistic style adaptions. Conversation partners thus adapt their communication style to become "more like their fellow interactants in a bid to decrease social distance, seek or signal approval, and thereby accommodate" (Giles et al. 2007, p. 142). By referring to an object, conversants develop a shared conceptualization or conceptual pact that they can refer to in subsequent discussions, resulting in lexical entrainment (Brennan and Clark 1996). In line with this definition, CAT suggests that people consciously or subconsciously accommodate their dialogue partners to develop closer relationships and signal empathy, credibility, and a common social identity (Giles et al. 2007). Across communication formats, conversants can accommodate one another by mimicking verbal behaviors, such as matching word choices and utterance length, or nonverbal behavior, such as facial expressions or posture (Ireland and Pennebaker 2010).

To derive linguistic styles and the proximity between two or more writing styles, researchers in communication science and linguistics highlight the use of *function words*, which indicate sentinel structure within texts (Tausczik and Pennebaker 2010). In contrast with nonfunction words (e.g., nouns, verbs, adjectives), which convey content, function words set the tone for social interactions and are key to understanding the relationships among speakers, objects, and other people (Chung and Pennebaker 2007). In line with CAT, the similar use of function words, and thus a greater LSM, represents a symbolic move toward stronger social identification and/or psychological synchrony (Ireland and Pennebaker 2010). Thus, LSM fosters understanding and decreases perceived social distance. For example, for couples on a first date, LSMs predict subsequent relationship viability; Huffaker et al. (2011) show that in online negotiations, greater matches in function word usage increase interpersonal rapport and agreement between potential coalition partners. Therefore, LSM unobtrusively signals conversant affinity and is diagnostic of behavioral outcomes regardless of the interaction environment, perceived quality, length, or objective (Niederhoffer and Pennebaker 2002).

Beyond an individual-level communicative alignment, verbal mimicry represents a key indicator for collective alignment (Kozlowski and Ilgen 2006) that can influence group performance (Brockman and Morgan 2006). Yet this communication style assimilation is distinct from so-called markers of community (Muniz and O'Guinn 2001), such as codified rituals and traditions, formal procedures, and isolated events. Communities may develop a distinctive collective communication style to achieve a sense of oneness or consciousness (Fayard and DeSanctis 2010). Members switch between communication styles to align or distance themselves from a particular collective (Gumperz and Levinson1996). Thus, we conceptualize an individual member's alignment with a common communicative style as a symbolic action, reflecting his or her identification with a community, which could affect subsequent participation efforts (Wasko and Faraj 2005).

In contrast with existing cross-sectional research, contemporary conceptualizations of members' socialization into groups consider it a dynamic, evolutionary process of alignment (Levine and Moreland 1994). For example, Elsbach and Bhattacharya (2001) demonstrate that employees fluctuate in their levels of identification with their organization. Similarly, Herring (2001) shows that people's communication patterns, rather than being stable, evolve over time. Therefore, different patterns of LSM within a community can emerge as a grouping principle and differentially signal members' identification with the collective. Previous research has shown that two properties of temporal development—the rate of change and the number of reversals in a trend—explain behavioral dynamics across scientific disciplines (DeKinder and Kohli 2008; Jokisaari and Nurmi 2009). Accordingly, it seems necessary to adopt a temporal dynamic perspective that focuses on interdependent communication incidents of members, across time and communities, to disentangle the parameters of temporal development in members' linguistic style. More precisely, we deduce LSM trends (degree of convergence toward or divergence from a collective style) and the number of reversals in these trends (changes from convergence to divergence, and vice versa) to probe their impact on members' participation quantity and quality in user communities. We develop hypotheses that reflect these relationships.

Hypotheses Development

Individual LSM

In user communities, we suggest that members' alterations in their linguistic style, to more closely match the community's dominant style, represent symbolic acts. They signal members' level of social identification and give rise to their subsequent participation efforts. Research on face-to-face interactions reveals conversants' tendency to accommodate one another by adapting verbal and nonverbal behaviors (Giles 2009); other research shows that in computer-mediated contexts, such communication style accommodations are manifest through interactants' assimilations in linguistic styles (Ireland and Pennebaker 2010). Research on communication accommodation further shows that LSMs elicit understanding and perceptions of a common social identity, but they decrease perceptions of social distance among partners (Giles et al. 2007).

Whereas the aforementioned studies exclusively examine communication dyads, the symbolic act integral to conversants' language use also may represent social identification in groups of conversants. This position is rooted in research on both organizational discourse (e.g. Hardy et al. 2005) and information systems (Fayard and DeSanctis 2010). In a study of online communities, Fayard and DeSanctis (2010), illustrate the impact of shared language use on members' feelings of "we-ness." They argue that the shared identification by members online is displayed through the enactment of a consistent discursive practice (e.g., common ways to phrase greetings and closings). Similarly, Postmes et al. (2005) show that participants in computer-mediated communication form group identity through shared communication norms such as common use of abbreviations or slang. The presence of such group identity may influence the importance that members attach to issues being discussed and their willingness to invest time and energy in addressing them (Hardy et al. 2005). Thus, the degree to which a member's linguistic style matches the overall community style should reflect his or her level of identification with the group and influence his or her participation.

> *H_1: A member's linguistic style match is positively related to subsequent (a) participation quantity and (b) participation quality.*

Individual LSM Patterns

Current literature on communication accommodation and LSM typically examines accommodation and its implications at one time or derives a single, aggregated accommodation level across all conversational intervals (Giles 2009; Ireland and Pennebaker 2010). Similarly, user community literature almost exclusively relates community perceptions to participation likelihood at a single point in time (Faraj et al. 2011). But rather than being passively formed and stationary, members' community identifications evolve over time and are characterized by constant and active negotiation and contestation within a particular collective (Elsbach and Bhattacharya 2001; Postmes et al. 2000). The importance of considering within-individual temporal patterns for predicting people's perceptions and behaviors has frequently been emphasized in communication (Chidambaram 1996; DeSanctis and Monge 1999) and organizational turnover (Kammeyer-Mueller et al. 2005) research. In particular, reducing temporal trajectories to mean-level differences obscures members' different adjustment trajectories, which can account for additional variance in member behavior (DeSanctis and Monge 1999). To establish how a user community grows on its new members, we consider the temporal development of new members' socialization processes, which foster their subsequent participation behavior (Koh et al. 2007; Levine and Moreland 1994). Drawing on the extended view of organizational identification (Elsbach and Bhattacharya 2001), we also assess whether differences in members' community identification development, as manifested in their LSM patterns, relate differentially to participation efforts.

Researchers denote three forms of identification development: identification, disidentification, and ambivalent/neutral identification (Elsbach and Bhattacharya 2001; Kreiner and Ashforth 2004). These forms constitute ways people can define themselves through attachments. In contrast with identification, disidentification acknowledges that members derive a sense of self by distancing (rather than aligning) themselves from a particular collective's attributes or principles (Elsbach and Bhattacharya 2001). Furthermore, members may have mixed, rather than one-sided, views regarding their fit with a collective. Ambivalent identification enables members to define themselves as the same as the collective at one time but different from it at other moments (Kreiner and Ashforth 2004).

LSM Trends

Disidentification consists of a pattern of increasing disconnect between aspects of the collective and a person's self-conception (Elsbach and Bhattacharya 2001) and behavior (Kreiner and Ashforth 2004). Considering the symbolic role of LSMs, such trends should become manifest themselves in members' linguistic style convergence or divergence with the community style. Beyond an average level, trends in members' LSMs can signal whether these members increasingly adapt the linguistic style of the community and contribute to the collective identity or whether they are departing from the established discursive practices and disregarding the (implicit) rules of engagement. In communication dyads, people who make adjustments in their communication style across conversational intervals display more interest in establishing common ground perceptions than conversants who do not (Ireland and Pennebaker 2010). Furthermore, recent research on the socialization of newcomers has highlighted that, beyond general trend considerations, studying the rate at which such trends occur yields finer-grained insights into temporal developments (Jokisaari and Nurmi 2009). Accordingly, we consider the rate of members' LSM trends, such that faster convergence (divergence) should signal greater identification (disidentification) and thereby increase (decrease) people's participation efforts.

> *H2: The greater the rate of convergence (divergence) in members' linguistic style matches, the higher (lower) their subsequent (a) participation quantity and (b) participation quality.*

LSM Reversals

In addition to solely identifying or disidentifying, members in a collective may be ambivalent about their identification (Elsbach and Bhattacharya 2001). Particularly in collectives that are rather loosely structured, members can simultaneously identify and disidentify with the organization (or aspects of it), maintaining an overall state of ambivalence (Kreiner and Ashforth 2004). Similarly, in user communities, members may be inclined to immerse themselves in some discussion topics but completely avoid others. Such mixed feelings cause members to feel torn between identification and disidentification (Kreiner and Ashforth 2004), which should lead them to reverse the direction of their LSM trend repeatedly. The total reversals in members' communication style thus reflect consistency (or lack thereof) in a member's identification behavior. Fayard and DeSanctis show that online community members who do not consistently participate in the "language game" (i.e., the shared communication style) are less likely to continue the expected behaviors. Similarly, Kreiner and Ashforth (2004) suggest that to the degree organization members experience ambivalent identification, they use up valuable cognitive resources that otherwise could be spent on organizational goals; they also appear reluctant to go beyond the required level of job performance. Thus, members with a relatively high degree of reversals in LSM development should be less motivated to continue to provide high-quality argumentation and participate.

> *H_3: Increasing amounts of reversals in members' LSM relate negatively to (a) participation quantity and (b) participation quality.*

Group-Level LSM

The impact of measures taken to stimulate members to actively contribute varies considerably across user communities (Iriberri and Leroy 2009). Although the drivers of success are many and often case specific (Faraj et al. 2011), active communities generally are characterized by synergistic social processes and shared understanding among members regarding their goals and behaviors (Moran and Gossieaux 2010). In practice groups, members' united behavior when working toward common goals (i.e., cohesiveness; Brockman and Morgan 2006) results in greater levels of interaction and shared focus among members, whereas differing patterns are detrimental to the overall group performance (Dennis et al. 2008). Specifically, cohesiveness in communicative behavior is associated with improved collaborations in online group settings, with reduced cognitive effort to encode and decode messages and thus improved response times in discussions (Kock 2004). The degree to which communicative behavior in user communities is "in sync" likely leads to differential success in stimulating overall community performance.

While all virtual communities exhibit discursive practices that provide scripts for action, there are also clear differences among communities. Fayard and DeSanctis describe online forums in which a small number of active members early on significantly shape the communication, creating the community atmosphere and encouraging a sense of we-ness

(Dholakia et al. 2004) . A distinct style emerges from initial interactions that allows other participants to join the forum, align themselves easily, and participate actively. In contrast, in communities without such established cohesive communication styles, it is difficult for members to identify and adopt the appropriate linguistic style (Fayard and DeSanctis 2010). Thus, the collaboration process is hampered, which also impedes members' ability to participate actively. User communities exhibiting high levels of LSM across members should reflect a more prominent collective identity, which affects member participation effort.

> H_4: *Community-level cohesiveness in members' LSM relates positively to individual members' (a) participation quantity and (b) participation quality.*

Method

Setting

The study sample includes 37 similarly structured user communities. Such communities "constitute an online social structure woven from continuous interactions among individuals focused around shared interests and common practices, as well as usage of the same tools and products" (Dahlander and Frederiksen 2012, p. 989). All of the user communities were hosted by the same market research consultancy, which aimed to facilitate cocreation with users spanning the following industries: finance and insurance (7 communities), information services (4), retail trade (7), manufacturing (10), arts and entertainment (6), and other services (3) between 2009 and 2011. We chose this setting for several reasons. First, members' community participation is, as with most user communities, purely intrinsically motivated, in that no financial rewards are given (Dholakia et al. 2004). Second, to motivate and sustain participation in such collectives, it is crucial for members to develop a sense of community identification (Kohler et al. 2011). Third, as with most online communities, nonverbal social cues and personal member information are not available, with text-based posts serving as the sole means by which to develop and assess social identification (Herring 2001). Fourth, the homogeneity across the user communities' setup, duration, structure, and purpose supports both within- and between-community comparisons. Each community consists of 150–300 members, all invited to participate because of their interest in and affiliation with the facilitating company and its products or services. Because this study's focus is on members' communication as symbolic action, we exclude all members who did not participate (post) from the sample (lurkers), which resulted in a final sample of 2,208 members across all 37 communities who made a total of 74,246 posts. Members were mainly male (56%) and on average were 37 years old. On average, there were 63 active members per community with 32 posts per member, and with a post length of 42 words across all communities. Members logged into the community 80 times and viewed 1,204 pages on average.

Data and Measures

A distinct language style in an online community develops in the first few interactions (Fayard and DeSanctis 2010). Therefore, we divided the observation period into an "initiation" period (T_1), encompassing a member's first two weeks in a community (Farzan et al. 2012), and an active participation period, covering the subsequent eight weeks of membership (T_2).

Dependent Variables

The absolute count of posts by each individual community member summed over the eight weeks in the active participation period, T_2, indicated the participation quantity per individual community member ($PQuant_i^{T_2}$):

$$PQuant_i^{T_2} = \sum_{t=1}^{8} Posts_{it}^{T_2} \tag{1}$$

where i stands for individual community member and *t* stands for week in the eight-week time period after the initial two-week initiation period (T_2).

Furthermore, because of the importance of substantiated and well-developed arguments in effective group discussions (Seibold and Meyers 2007), we consider argumentative development, generically conceptualized as quality, in posts. To construct this measure, we followed Cohn et al. (2004) and text-mined all community posts across the 37 communities to determine the use of causal words (e.g., *because*, *cause*, *effect*) and other words suggestive of cognitive processing (e.g., *realize*, *understand*). We constructed a composite measure of the text-mined cognitive effort ($CE_{ip}^{T_2}$) for each individual (i) by post (p) by summing the total amount of cognitive words and causal words used in a post, where T_2 *indcates* the time period after the initial two-week initiation period:

$$CE_{ip}^{T_2} = CognitiveWords_{ip}^{T_2} + CausalWords_{ip}^{T_2} \tag{2}$$

To verify the accuracy of this approach for capturing argument development, we drew a random subset of 6,000 community posts. Three independent coders content analyzed and classified the subset of posts according to the quality of how insights are shared using the thought-listing technique commonly employed to assess the quality of cognitive responses (Cacioppo and Petty 1981). In line with prior research using this technique, the coders considered every stated and reasoned idea, regardless of whether they were grammatically correct, as a unit. The intercoder agreement was assessed using Krippendorff's α, which found matching argument count incidences of 93 percent (well above the critical threshold of .80). To verify the level of congruence between the count of arguments by the three manual coders and the text-mined cognitive effort per post, we computed a Pearson product–moment correlation coefficient. The correlation coefficient of .72 prompts strong confidence in the viability of the text mining approach. The average, text-mined count of cognitive effort words per post during period T_2 calculated for each user community member yielded the second dependent variable, namely, the average argument development quality an individual member exerted in his or her posts ($PQual_i^{T_2}$):

$$PQual_i^{T_2} = \sum_p \frac{CE_{ip}^{T_2}}{PQuant_i^{T_2}} \tag{3}$$

where i is the individual, p refers to post, T_2 denotes the time period after initiation, and $CE_{ip}^{T_2}$ is the cognitive effort per post, as derived in Equation (2).

Independent Variables

As mentioned previously, a distinct language style in an online forum develops during the first few interactions (Fayard and DeSanctis 2010). Therefore, we assessed the independent variables in this study according to communication behaviors in the first two weeks (T_1) only. Observing communication during the initiation period T_1 and linking it to participation behavior during the observation period T_2 also ensures greater causality implications, because the predictor variables precede the outcome behaviors. In line with recent research on LSM (Gonzales et al. 2010; Ireland and Pennebaker 2010), we operationalized LSM as a measure of the degree to which two or more conversants produce similar usage intensities of function words.

First, we text-mined for each user community post the total word count of nine function word categories (which comprise all 469 function words in English): (1) auxiliary verbs (e.g., *to be*, *to have*), (2) articles (e.g., *an*, *the*), (3) common adverbs (e.g., *hardly*, *often*), (4) personal pronouns (e.g., *I*, *they*, *we*), (5) impersonal pronouns (e.g., *it*, *those*), (6) prepositions (e.g., *for*, *after*, *with*), (7) negations (e.g., *not*, *never*), (8) conjunctions (e.g., *and*, *but*), and (9) quantifiers (e.g., *many*, *few*) (Ireland and Pennebaker 2010). We constructed a measure of the function word usage intensity ($FWC_{ij}^{T_1}$) for each individual community member (i) and for each function word category (j) by dividing the total number of words belonging to the particular function word category across all posts by the total number of words per the post (p) across all posts:

$$FWC_{ij}^{T_1} = \frac{\sum_p FWC_{ijp}^{T_1}}{\sum_p TotalWords_{ip}^{T_1}} \tag{4}$$

Second, we determined the specific community-level function word category usage intensity ($FWC_{cj}^{T_1}$) for each community (c) and for each category (j) across all the posts that were present before the current post was submitted. That is, to derive the community's usage style of a particular function word category in period T_1, we took the average of that function word category's usage intensity across all posts in the given community made in T_1:

$$FWC_{cj}^{T_1} = \sum_{j \in c} \frac{FWC_{ij}^{T_1}}{PQuant_i^{T_1}} \tag{5}$$

Third, language style match (LSM_{ij}) for each individual community member (i) for each function word category (j) is derived using the following equation:

$$LSM_{ij} = 1 - \frac{\left(|FWC_{ij}^{T_1} - FWC_{cj}^{T_1}|\right)}{\left(|FWC_{ij}^{T_1} + FWC_{cj}^{T_1} + 0.0001|\right)} \tag{6}$$

In this equation, LSM_{ij} is the ratio of overlap between the usage intensity by each individual member (i) for each function word category (j) and the cumulative average usage intensity of the same function word category (j) by all community posts that were posted prior to the current post in the community (c). In the denominator, we added .0001 to prevent empty sets. Finally, we established the overall LSM for each post across all nine functional word categories by averaging the nine separate LSM ratios for each function word category.

We used two levels to describe the dynamics of communicative behaviors: member-level aspects of communicative behavior (average, trend, and reversals) and overall community-level cohesiveness in communicative behavior. Following our theoretical development, each of these parameters highlights unique aspects of members' communicative trajectories in the sample. First, the within-member LSM mean is the general level of LSM across all posts of member i in the first two weeks (T_1). It is computed by taking the mean LSM across all posts of member i in T_1, which yields a composite LSM ratio, bounded by 0 and 1 for each of the 2,261 members, for which higher numbers represent greater overall stylistic similarity between a member and the community in T_1 (see Gonzales et al. 2010). We therefore use it to signify a member-specific average level of LSM across all posts in the first two weeks (T_1).

Second, a member's LSM trend is established by regressing the sequential post incidences by member (i) on the respective LSM of each of his or her posts using the least squares method. All post incidences of a member are numbered consecutively, so that posts made later receive higher numbers. We then regressed the LSM of member i on the post incidents of member i. The beta-coefficient (β_1) of the post incidence variable in the equation $LSM_i = \beta_0 + \beta_1 \cdot PostIncidents_i + \varepsilon_i$ for member i signifies the rate of change in his or her LSM trend over time (Jokisaari and Nurmi 2009). A rate of change near zero represents a stable LSM trend throughout the initial period of a person's community membership. The more positive the coefficient, the greater the assimilations in linguistic style by that member.

Third, we measured the frequency of change in a member's LSM across subsequent posts by counting the number of slope changes (that exceed one standard deviation) compared with the number of posts (DeKinder and Kohli 2008). We did not count slope changes smaller than one standard deviation, because they only reflect minor alterations in an otherwise stable communication style.

Fourth, to establish community-level cohesiveness in communicative style (i.e., indicate the degree of equality in communication styles within a community), we used the coefficient of variation adjusted for group size, as suggested by Harrison and Klein (2007). The majority of research on collective settings has examined group behavior similarities by using direct consensus models (i.e., taking the group average as the preferred mode of aggregation). Yet recent research shows that such group-level similarities are more appropriately assessed by considering within-group variability or dispersion-composition models, which better reflect multilevel phenomena (Cole et al. 2011). In recognition of these new insights, we constructed the measure of cohesiveness in communication style within communities as one minus the within-group variability of LSM.

To empirically justify this multilevel perspective and validate the aggregation procedure, we calculated the within-group agreement measure r_{wg} for single-item measures, as James et al. (1984) suggest. Because this measure is designed for scale variables only, we discretionized the original ratio measure of LSM into a scale consisting of 10 categories and estimated the rectangular distribution. The mean of the r_{wg} coefficients, which indicates the homogeneity of LSM within communities, is .87. These findings demonstrate that the common communication style of individual LSMs within communities is highly consistent. We illustrate the individual-level LSM parameters using a community discussion excerpt from one of the user communities in the data set (Table 1).

Multiple users commented and exchanged recipes and tips for cooking and baking; for purposes of illustration, we focus on User 3. Although the posted content varies, users' expressions (linguistic style) are relatively similar, with a warm and personal tone. Consider User 3: Her LSM scores are .12, .56, and .34 across her three posts, compared against the cumulative communication style of the overall community (which includes more posts than listed here). Her second post matched the community linguistic style best in terms of function words. Her first post is highly divergent from the conventional communication style in that user community. Note further that (considering only these three posts), User 3's overall rate of change would be negative, and her linguistic style would drift away from the community style. Finally, across these three post incidences, one change (reversal) appears in User 3's LSM trend; that is, there is a trend toward a greater LSM from her first to her second post, but this trend reverses with her third post.

Control Variables

In addition to communication style, several member- and group-related aspects of user communities affect members' participation quantity and quality. Members' participation history can help predict their future participation (Moran and Gossieaux 2010). To capture differences in members' general disposition to participate in a user community, we controlled for their participation quantity ($PQuant_i^{T_1}$), participation quality ($PQual_i^{T_1}$), and number of page views within the community ($PViews_i^{T_1}$) during their first two weeks of membership (T_1). Previous research has suggested that all these variables influence participation efforts (Koh et al. 2007; Ma and Agarwal 2007). Because participation quality is a count measure, we needed to control for the average word length ($PLength_i^{T_1}$) of a member's posts as well. Furthermore, at the community level, we controlled for the community size ($CSize_c$), which is the total amount of members. Moreover, we controlled for community-level qualities, which may affect the influence of community members' identification on their participation. Specifically, we consider overall participation quantity ($CPQuant_c^{T_1}$) and quality ($CPQual_c^{T_1}$) by all members of a particular community. Table 2 outlines the descriptive statistics and correlations; the correlations be-

Table 1. Sample Conversations in a User Community: Recipe Discussions

Member	Text	LSM of Member
User 1:	Hi, my family love cake - particularly a nice light victoria sponge. I use …, it makes the cake really moist. My Mum taught me this when I was a nipper.	0.69
User 2:	Well me and my hubby used to go eat in … I kept thinking i used to make that many years ago, so i set about it didn't tell hubby. Just put at front of him, well he loved it, said it is better than … its good but fattening	0.71
User 3:	Lost weight at … and devised a lot of convenient low fat / fat free recipes when I was on the plan. The ideal way to make … is to use …, season appropriately, no fat other than the substances in the milk.	0.32
User 2:	Thanks … I am going to look out for some of that - I do like … but I have to lose quite some weight so this is just the job! Have you discovered … It really does fill me up honest and it tastes wonderful too!	0.57
	…	
User 4:	I LOVE … Hollondaise sauce but you can't buy it in England. I stocked up big time when I was working in the US but it's all gone now - all gobble up and my family are suffering withdrawl!! …. Now I need to make my mock hollandaise every time we have cauliflower with our dinner!	0.73
User 5:	Hi … if you like Mock Hollandaise sauce for Cauliflower without the … this one may be for you. Just use any White sauce recipe and add the nutmeg, lemon juice and egg yolk this makes all the difference and done really quickly.	0.62
User 3:	Talking about …, I also like the range of …. I use the ITALIAN GARLIC one all the time, so easy to shake and add … using this one mixed up with butter is very easy to do and is ready in a flash.	0.56
	…	
User 3:	When roasting a joint of beef use …. Learned this years ago working as a waitress at a night club that had a top notch restaurant. The chef was Croatian and his food was devine!	0.34
User 6:	Hi everyone, for an interesting twist I have a tip for you, use peanut butter as a topping on a beefburger. It sounds odd but tastes really good! I saw this on … and thought I'd give it a try and have used up many jars of peanut butter since!	0.72

Note: These excerpts are taken from one of the user communities in the data set. Multiple users commented and exchanged recipes and tips for cooking and baking. For privacy reasons, we removed user names, brands, and product names and parts of the texts.

Table 2. Descriptive Statistics and Correlations[a]

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
Member-Level Variables															
1. PQuant^{T2}	14.79	23.86	1.00												
2. PQual^{T2}	1.84	0.05	0.02	1.00											
3. LSMM^{T1}	0.63	0.17	0.35	0.25	1.00										
4. LSMT^{T1}	-0.02	0.09	0.12	0.09	0.01	1.00									
5. LSMR^{T1}	0.27	0.28	-0.16	-0.14	-0.15	-0.03	1.00								
6. PQuant^{T1}	10.19	14.66	0.47	0.43	0.10	0.09	-0.14	1.00							
7. PQual^{T1}	0.05	0.05	-0.03	0.08	0.18	0.01	-0.24	0.01	1.00						
8. PView^{T1}	615.00	223.39	0.26	0.20	-0.02	0.01	-0.03	-0.03	-0.05	1.00					
9. PLength^{T1}	41.99	10.48	-0.06	0.29	0.29	-0.01	-0.29	-0.05	0.20	0.03	1.00				
Community-Level Variables															
10. CLSM^{T1}	0.91	0.15	0.14	0.27	0.17	0.02	0.02	0.18	0.13	-0.03	-0.01	1.00			
11. CSize	178	33.43	0.12	0.11	-0.02	-0.08	0.10	-0.06	-0.06	-0.01	-0.08	0.19	1.00		
12. CPQuant^{T1}	9.03	6.44	0.03	0.01	0.13	-0.01	0.13	0.42	-0.01	0.03	-0.01	0.15	-0.09	1.00	
13. CPQual^{T1}	0.05	0.04	-0.12	0.62	0.01	-0.02	-0.03	-0.07	0.54	-0.01	0.21	0.14	-0.24	-0.01	1.00

[a]n = 2208

tween participation quantity, participation quality, and our communication style variables of interest are all in the expected directions.

Prestudy

We conducted a prestudy to assess the validity of LSM empirically as a symbol for members' social identification. Of the five 7-point Likert scales that Algesheimer et al. (2005) developed to measure community identification, we used the four that were most applicable in this context: "I am very attached to the community," "other community members and I share the same objectives," "I see myself as a part of the community," and "the friendships I have with other community members mean a lot to me." We sent online surveys twice (including one follow-up survey) by e-mail to all members of a subset of 18 communities 8 to 12 weeks after they had begun their community participation. We introduced the questions as part of a general opinion survey regarding their community experience. Of the 3,211 community members, 622 completed the survey (response rate = 19.37%). The scale items were internally consistent (coefficient α = .74). We considered the correlation between members' community identification on their overall degree of LSM (established from their posts). The results of the Pearson–product correlation coefficient showed a significant correlation between members' LSM ratio and their community identification (.71, $p < .01$). On the basis of this prestudy result, we are confident that the degree of members' LSM is a significant positive symbol of their community identification.

Data Analysis

To capture the influence of the explanatory variables at the member and community levels on members' subsequent behavior, we specified two multilevel Poisson models, often referred to as hierarchical linear models (HLMs). The Poisson HLM approach accounts for member interdependencies and simultaneously allows for investigations of cross-level effects (Long 1997). With multiple members nested in each user community, the HLM modeling approach also controls appropriately for the possibility that communication behaviors from members in the same community may be more similar than they are for members in another community; it can estimate relationships that are nested across levels. We estimated how much variance in members' participation quantity and argument quality resides within members or between communities by computing the median incidence rate ratios (IRRs) (Long 1997). With regard to participation quantity, the estimated median IRR is 1.45, which implies that half of the time, the ratio of expected participation quantity will range from .68 (1/1.45 = .68) to 1.45, and the other half of the time, it will lie outside that range. Similarly, the median IRR for participation quality is estimated at 1.38, and thus, the expectation should lie in the range between .73 and 1.38 in 50 percent of the cases. This finding provides convincing evidence that community characteristics can have a direct influence on members' participation quantity and argument quality. We next specified two multilevel Poisson regression models to estimate the effects of the antecedent member- and community-level variables at T_1 (and time-fixed variables) on participation quantity and quality at T_2. We used STATA12 to estimate the models, beginning with two null (intercept only) models for participation quantity and argument quality. We introduced the individual-level variables and covariates in Model 1a for participation quantity and Model 1b for argument quality. Finally, we added the group-level variables and covariates to estimate the full models for participation quantity (Model 2a)

$$\begin{aligned} PQuant_{ic}^{T_2} = \exp(&\beta_{0c} + \beta_1 \cdot LSMM_{ic}^{T_1} + \beta_2 \cdot LSMT_{ic}^{T_1} + \\ &\beta_3 \cdot LSMR_{ic}^{T_1} + \beta_4 \cdot LSMC_{c}^{T_1} + \beta_5 \cdot PQuant_{ic}^{T_1} + \\ &\beta_6 \cdot PQual_{ic}^{T_1} + \beta_7 \cdot PViews_{ic}^{T_1} + \beta_8 \cdot PLength_{ic}^{T_1} + \\ &\beta_9 \cdot CSize_c + \beta_{10} \cdot CPQuant_c^{T_1} + \beta_{11} \cdot CPQual_c^{T_1} + \zeta_{0c}) \end{aligned} \tag{7}$$

and participation quality (Model 2b)

$$\begin{aligned} PQual_{ic}^{T_2} = \exp(&\beta_{0c} + \beta_1 \cdot LSMM_{ic}^{T_1} + \beta_2 \cdot LSMR_{ic}^{T_1} + \\ &\beta_4 \cdot LSMC_{c}^{T_1} + \beta_5 \cdot PQuant_{ic}^{T_1} + \beta_6 \cdot PQual_{ic}^{T_1} + \beta_7 \cdot \\ &PViews_{ic}^{T_1} + \beta_8 \cdot PLength_{ic}^{T_1} + \beta_9 \cdot CSize_c + \beta_{10} \cdot \\ &CPQuant_c^{T_1} + \beta_{11} \cdot CPQual_c^{T_1} + \zeta_{0c}) \end{aligned} \tag{8}$$

where

- i is the individual member, c indicates the community, and $PQuant_{ic}^{T_2}$ and $PQual_{ic}^{T_2}$ stand for the participation quantity and participation quality, respectively, in period T_2.
- For the hypothesized effects, $LSMM_{ic}^{T_1}$ is the average degree of LSM, $LSMT_{ic}^{T_1}$ is the rate of change in LSM trend, $LSMR_{ic}^{T_1}$ is the frequency of reversals in LSM in a member's posts in period T_1 and $LSMC_{c}^{T_1}$ is the group-level cohesiveness in members' LSM at the community level in period T_1.
- The covariates at the individual member level are participation quantity ($PQuant_{ic}^{T_1}$), participation quality

($PQual_{ic}^{T_1}$), page views ($PViews_{ic}^{T_1}$), and average post length ($PLength_{ic}^{T_1}$). At the community level, we control for community size ($CSize_c$), overall participation quantity ($CPQuant_c^{T_1}$), and overall participation quality ($CPQual_c^{T_1}$) of the community.

- ζ_{0c} is the random variance for intercept.

We assume an independent correlation matrix. The correlation matrix in Table 2 and the variance inflation factor scores indicate that there is no potential threat of multicollinearity (Model 2a $VIF_{MAX} = 1.24$; Model 2b $VIF_{MAX} = 1.97$). To improve interpretability, we standardized all predictor variables in the models.

We used full maximum likelihood estimation to estimate the parameters and compare the model fits across nested models (Raudenbush and Bryk 2002). Using χ^2 difference tests, we confirmed that the member- and community-level explanatory variables added explanatory power to the final model (see Table 3, Models 2a and 2b). We took all of the estimates analyzed next from these final models; the parameter estimates provide support for the majority of the hypotheses. Finally, the incidence rate ratio (IRR) can be used as an effect size measure for multilevel Poisson regression models and indicates the percentage change in the independent variable for one unit of change in the dependent variable.

The results reveal a significant, positive impact of members' LSM in the early period (T_1) on their community membership participation (${\beta_{LSMM}}^{T_1} = .23, p < .001$; quality ${\beta_{LSMM}}^{T_1} = .33$, $p < .001$) in the subsequent membership period (T_2). Thus, we can confirm H_1. The IRRs indicate that a one standard deviation increase in LSMM is associated with increases of 26 percent (IRR = 1.26, CI (95%) 1.22 – 1.29) in participation quantity and 39 percent (IRR = 1.39, CI (95%) 1.29 – 1.49) in participation quality. Both shape parameters, capturing the temporal development in linguistic style over members' initial membership period, have significant effects on subsequent participation quantity. A faster accommodation rate by members toward a stronger LSM in T_1 has a significant positive effect on participation quantity (${\beta_{LSMT}}^{T_1} = .06, p < .001$) and quality (${\beta_{LSMT}}^{T_1} = .17$, $p < .001$) at T_2, in support of H_2. Conversely, reversals and frequent alterations in members' LSMs toward the community style have a significant negative impact on their participation quantity (${\beta_{LSMR}}^{T_1} = -.07$, $p < .001$) but no significant impact on argument quality at T_2, in support of H_{3a} but not H_{3b}. Overall, the results support the claim that the temporal development of members' communication style can explain subsequent behavior. A trend increase by one standard deviation correlates to a 6 percent increase in participation quantity and an 18 percent increase in quality. A one standard deviation increase in LSM reversals decreases participation quantity by 7 percent. Furthermore, in support of H_4, the community-level effect of communication style cohesiveness in communities in T_1 has a significant positive impact on both participation quantity (${\beta_{LSMC}}^{T_1} = 1.05, p < .001$) and argument quality (${\beta_{LSMC}}^{T_1} = .46$, $p < .001$). As the IRR indicates, a one standard deviation increase in the cohesiveness of the communities' linguistic style (i.e., all members use a more synchronized linguistic style) enhances individual members' participation quantity by 147% and their participation quality by 103% (Long 1997). This highlights the criticality of community-level context aspects for members' behaviors. Table 4 provides an overview of all the results. The member-level control variables are all significant predictors of subsequent participation quantity, but only members' participation quality in T_1 significantly affects participation quality in T_2 (${\beta_{PQual}}^{T_1} = .23, p < .001$). None of the community-level covariates significantly affects members' subsequent participation quantity. However, community size significantly reduces members' subsequent participation quality ($\beta_{CSize} = -.27, p < .01$), and a community in which all members develop their arguments more in their posts significantly and positively influences the subsequent participation quality of the individual members (${\beta_{CPQual}}^{T_1} = -.36, p < .001$).

Discussion and Conclusion

Theoretical Implications

This study conceptualizes a user community member's LSM with the community's common communication style as symbolic action. In line with contemporary views on the role of language (e.g., Herring 2001), this study posits and empirically validates how linguistics serves not only to describe realities but also to symbolically reflect conversants' feelings and affect their subsequent behaviors. The symbolic action innate to communication style is the foundation of speech act theory (Searle 1975), rhetoric (Gill and Whedbee 1997), and social constructionism (Berger and Luckmann 1966). However, a relative paucity of research describes the signaling role of linguistic styles online (Fayard and DeSanctis 2010), which entail no nonverbal cues and thus distinct communication behaviors. Focusing on communication enriches extant research on drivers of community participation that has focused on members' motives (e.g., Koh et al. 2007), expertise (e.g., Wasko and Faraj 2005), or network position (e.g., Dahlander and Frederiksen 2012) as well as community designs (e.g., Ma and Agarwal 2007).

Table 3. Multilevel Poisson Regression Analysis

	Participation Quantity^{T2}			Participation Quality^{T2}		
Constructs	**Model 1a**	**Model 2a**	**IRR (95% CI)[a]**	**Model 1b**	**Model 2b**	**IRR (95% CI)[a]**
Member-Level Variables						
LSMM^{T1}	0.71**	0.23**	1.26 (1.23–1.30)	0.63**	0.33**	1.39 (1.29–1.50)
LSMT^{T1}	0.09**	0.06**	1.06 (1.05–1.08)	0.20**	0.17**	1.19 (1.14–1.24)
LSMR^{T1}	–0.04**	–0.07**	0.93 (0.97–0.95)	–0.02	–0.01	0.98 (0.94–1.03)
PQuant^{T1}	0.35**	0.31**	1.36 (1.35–1.37)	0.02	–0.03	0.97 (0.93–1.01)
PQual^{T1}	0.08**	0.07**	1.07 (1.04–1.10)	0.27**	0.23**	1.26 (1.19–1.34)
PView^{T1}	0.12**	0.11**	1.12 (1.11–1.13)	–0.02	–0.03	0.96 (0.92–1.01)
PLength^{T1}	–0.02**	–0.01**	0.99 (0.98–0.99)	0.01**	0.02**	1.02 (1.01–1.02)
Community-Level Variables						
CLSM^{T1}		0.90**	2.48 (2.39–2.56)		0.72**	2.03 (1.84–2.29)
CSize		–0.20	0.81 (0.63–1.06)		–0.27**	0.74 (0.62–0.94)
CPQuant^{T1}		–0.19	0.83 (0.67–1.01)		–0.10	0.91 (0.77–1.07)
CPQual^{T1}		0.11	1.05 (0.84–1.32)		0.36**	1.74 (1.46–2.11)
Intercept	3.05**	2.54**		–0.17	–0.38**	
N (members)	2208	2208		2208	2208	
N (communities)	37	37		37	37	
Log likelihood	–13498.30	–12104.40		–3042.21	–2932.29	
Wald χ^2 (df)	19634.35(7)	21486.63** (11)		1231.19 (7)	1376.78 (11)	
Deviance(–2LL)[b]	26996.59	24208.79		6084.43	5864.58	

*p < .05 and **p < .01.
[a]IRR = incidence rate ratio; CI= confidence interval
[b]For Models 2a and 2b, the LR test is significant (p < .001), indicating a relative increase in model fit compared with Models 1a and 1b, respectively.

Table 4. Hypotheses Testing

Hypothesis	**Relationships[a]**	**Direction**	**ß**	**Result**
H_{1a}	LSM Mean$_{(t1)}$ → Participation Quantity$_{(t2)}$	Positive	.23**	Supported
H_{1b}	LSM Mean$_{(t1)}$ → Participation Quality$_{(t2)}$	Positive	.33**	Supported
H_{2a}	LSM Trend$_{(t1)}$ → Participation Quantity$_{(t2)}$	Positive	.09**	Supported
H_{2b}	LSM Trend$_{(t1)}$ → Participation Quality$_{(t2)}$	Positive	.17**	Supported
H_{3a}	LSM Reversals$_{(t1)}$ → Participation Quantity$_{(t2)}$	Negative	-.04**	Supported
H_{3b}	LSM Reversals$_{(t1)}$ → Participation Quality$_{(t2)}$	Negative	n.s.	Not Supported
H_{4a}	LSM Cohesiveness$_{(t1)}$ → Participation Quantity$_{(t2)}$	Positive	.90**	Supported
H_{4b}	LSM Cohesiveness$_{(t1)}$ → Participation Quality$_{(t2)}$	Positive	.72**	Supported
Fit Measures	**Outcome Construct**	**Model**		**N(members)**
LL	Participation Quantity	–12104.40		2208
	Participation Quality	–2932.29		2208

*p < .05; **p < .01; n.s. = not significant.
Part. Quantity$_{(t2)}$ = Member's Participation Quantity in the second time period, Participation Quality$_{(t2)}$ = Member's Participation Quality in the second time period.

First, inspired by communication accommodation research in conversation dyads (Giles 2009; Ireland and Pennebaker 2010), this study contributes to contemporary research on the role of communication in user communities by showing that the degree of members' LSM symbolically reflects their level of identification in online groups (i.e., user communities). In doing so, our findings extend research on CAT to a multilateral setting by demonstrating its multilevel implications for members' participation. Previous research has shown that members' community identification critically influences their participation efforts (Kozlowski and Ilgen 2006; Wasko and Faraj 2005), yet it has not considered how text-based communication may drive participation efforts, rather than being just its result (Fayard and DeSanctis 2010). Such a language-as-action perspective diverges from a traditional language-as-product view (Brennan and Clark 1996) and advances language beyond description to become a means to construct reality.

Moreover, emerging research on CAT indicates that LSM is a subconscious, coordinative indicator of common ground formation, based on a sense of belonging to a community, yet might be beyond the members' control. In contrast, the traditional view highlights common ground formation as a cognitive, intentional attempt by people to replicate mental models to ease their understanding and facilitate shared sense-making (Gonzales et al. 2010; Weick 1995). Although easing communication might be an important rationale for group alignment, the results indicate a member's identification is the mechanism underlying LSM. The LSM's focus on function words enriches the conventional view on text analysis, which assumes that the semantics of nouns and verbs are key in understanding text. Instead our study identifies the use of function words as a subtle, implicit way of rendering other symbols as meaningful and interpretable for community participants.

Second, research on communication accommodation has focused on conversants' perceptions and behaviors at a given time but neglected the dynamic nature of a person's identification with a group (Levine and Moreland 1994). The current study extends CAT by adopting a temporal perspective, viewing LSM trends and reversals as essential, distinct symbols in the ongoing socialization process through which user community members produce, reproduce, and change their community identification. In line with the expanded view of organizational identification (Elsbach and Bhattacharya 2001; Kreiner and Ashforth 2004), steeper trends in linguistic style alignment (distancing) signal members' accelerated identification (disidentification) with the user community. Accordingly, the findings show that positive, faster trends toward greater LSM enhance members' subsequent participation quantity and quality. Frequent changes in linguistic style accommodation trends suggest members' ambivalence regarding their identification with a community (Elsbach and Bhattacharya 2001). In line with research into such identification uncertainty (Kreiner and Ashforth 2004), the current study shows that members who often alter their degree of LSM are less likely to excel in subsequent participation quantity. Yet contrary to the hypothesized relationship, reversals in members' LSM are not significantly negatively related to their subsequent participation quality. We speculate that the hypothesized negative impact of ambivalent identification on the willingness to interact with other members (Meyerson and Scully 1995) might be offset by its advantages. As Stonequist (1937, p.155) argues, members with ambivalent identifications are fundamentally "outsiders within," who access the knowledge of an insider but hold the critical attitude of the outsider. Thus, their overall participation quantity decreases, but the quality of their argument development remains unaffected, with these members serving as acute, able critics (Meyerson and Scully 1995). Overall, we demonstrate empirically the substantive symbolic nature of two temporal parameters (trend and reversals) in CAT, while controlling for behavioral (e.g., passive reading behavior) and contextual (e.g., community quality, size) aspects (Faraj et al. 2011; Ma and Agarwal 2007).

Third, we contribute to research on communication by considering text not only as a data source but also as the societal context in which the text is nested (Fairclough 1992). We find that synchronicity in communicative behavior—or cohesiveness across community members' linguistic styles—adds substantively to the explanation of individual members' participation behavior. Like other forms of cohesiveness (e.g., shared interpretative schemes), synchronous communication styles in groups appears to foster shared identification or a sense of we-ness among members, which further encourages members to invest themselves on behalf of the collective (Hardy et al. 2005). The significant explanatory power of the community-level characteristics (i.e., communication style, size, and participation quality and quality), which collectively explain 41 percent of variation in subsequent individual members' behavior, demonstrate the importance of contextualizing linguistics suggested in previous research (Fairclough 1992).

In corroboration with previous research, the current study offers additional insights on user community participation. Specifically, whereas previous research has stressed the criticality of frequent participation and the generation of good quality content (Ransbotham and Kane 2011), drawing on research on argumentation quality, we highlight the importance of argument development quality in group communication processes (Seibold et al. 2010). Well-formulated and well- developed argumentation enhances the effectiveness and value of members' community participation (Hansen and Haas 2001; Ransbotham and Kane 2011), possibly even to the point of transcending the importance of the actual content strength of the argument itself (Seibold et al. 2010). Accordingly, we view participation quality as the degree to which members substantiate and develop their statements and arguments, which enables a better analysis of participation quality across communities that vary in their content.

To validate the quality measure and to demonstrate its content independence, we drew on existing research on the conversational argument coding scheme (CACS), which has proved its utility and validity in assessing interpersonal arguments (Seibold et al. 2010; Seibold and Meyers 2007). Among other content categories, the scheme includes diametrically opposite content categories, such as convergence-seeking activities (i.e., showing agreement and recognition) and disagreement-relevant intrusions (i.e., showing denial and questioning accuracy and truth).

We drew a random subset of 32 community discussion threads (979 community posts). Two trained coders independently coded each post using the CACS. The Pearson product–moment correlation with our text-mined participation quality and presence of agreement/convergence-seeking behavior (.691, $p < .001$) and disagreement-relevant intrusions (.549, $p < .001$) are positive and significant, indicating the measure's validity across content categories. This conceptualization might also explain the seemingly contradictory findings of previous research, which suggest that group cohesiveness impedes performance. For example, Brockman and Morgan (2006) note a negative influence of group cohesiveness, yet they conceptualize performance as the development of novel and different ideas.

Limitations and Further Research

As does any form of academic inquiry, the current study suffers from several limitations, which can be used to identify fruitful lines for research. First, we derived the measure of linguistic style from a member's particular usage of functional words. Although this approach parallels previous research in linguistics and communication science, it does not account for the occurrence and impact of particular, community-specific vocabulary. There may be slang, positive and negative affective words, or other nonfunctional words that are strongly embedded in the communication style of a particular community (Postmes et al. 2005) and predict participation even stronger. Therefore, depending on availability of innovations in text analytic software, researchers may be able to capture the use of such style formats and assess their implications in member posts.

Second, we consider and empirically validate within-community congruence with one unique communication style. Although this approach is feasible for relatively small-scale user communities (not exceeding 200 active members), it may be more complicated to generalize these findings to larger or multifocused online communities, established around brands or for the purpose of peer-to-peer, after-sales support. For example, in such multifaceted environments, subgroups may form sub-communication styles, which may differ from the dominant style and perhaps indicate factionalism or community decline. Further research might investigate the value of communication styles as signals of subgroups within a single collective and relate this prediction to the life span of the community. Such an analysis would be useful for examining whether and how LSM analysis can be used in conjunction with existing modes of social network analysis.

Third, given the multilevel nature of our data, the accompanying sample size, and the use of Poisson regression models, we need to be aware that statistical and practical significance cannot be equated (Straub and Ang 2011). This study addresses a real-life phenomenon, holds statistically significant results for practically relevant parameters, and its findings may influence community managers' behaviors as spelled out below. However, even when an effect is significant and its IRR indicates a strong impact, achieving an improvement in the particular predictor variable to benefit participation quantity and quality may prove difficult to achieve, especially regarding community-level variables.

Fourth, this research focuses on members' socialization processes, which typically manifest themselves within the first weeks of participation (Farzan et al. 2012). Although we split member participation into this typical initiation period and the subsequent period of membership, there may be further important membership life cycle stages to be investigated.

Finally, this study considers communication style aspects alone as symbolic for members' social integration into a user community. Social integration signifies members' willingness to invest on behalf of the user community, but multiple case studies also recognize that members can be intentionally counterproductive or abolish proactive participation in communities (i.e., flaming). Such "flame wars" escalate over time and lead to detrimental effects on collaboration. Although flaming can stem from a variety of issues, including misunderstanding, frustration, or perceptions of unfairness, further research should seek text analytic approaches to detect flaming before it becomes detrimental to community collaboration.

Managerial Implications

This study's findings offer several actionable implications for managers of user communities. Overall, user-community managers, who struggle to encourage and sustain members' participation efforts, should recognize that community posts may reveal more about a member's attitude and perception than what is literally said. By introducing LSM and assessments of temporal development, the current study offers managers an unobtrusive and automatic way to assess user community members' social identification and predict their subsequent participation, provided members give their consent. In the absence of extrinsic remuneration, social identification offers a primary, intrinsic motivator of ongoing participation efforts. Four key implications for practitioners thus follow from our study.

First, to assess new members' integration, community managers should derive and track the dominant communication style by text mining the common usage style of function words. Firms can install continuous monitoring tools to assess not only the overall level of members' LSMs but also the patterns of how members' communication styles evolve through their iterative posting incidences. With this information, a community manager can obtain a real-time update of the sense of belonging by each member and detect changes before they lower participation behavior.

As Joëlla Marsman (2012), Insight Manager at the food company H. J. Heinz, states,

> Monitoring resemblance in writing style of our community members could really help us, as community managers, to prevent unnecessary unhappiness and drop out. However, to me the most important benefit is the opportunity it provides to boost participation and consequently collect richer data.

In many cases, a simple "thank you" or other symbolic acknowledgment at the right time may revive participation efforts.

Second, managers could leverage the emergence of community-level cohesiveness in communication style to drive individual members' subsequent participation efforts. By monitoring the variability in members' communication styles, managers can assess the strength of their connection. When styles begin to diverge, it may be necessary to host community events, such as brandfests, to spark a feeling of we-ness (Algesheimer et al. 2005) and encourage members to participate in their community.

Third, companies often decide to sponsor or buy into an existing user community on the basis of the number of community members or posts (Ransbotham and Kane 2011). However, they might more fruitfully assess the cohesiveness in the communication style across members to estimate the level of participation they can expect in the future.

Fourth, the amount of cognitive words used in community posts is a reliable measure of the amount of in-depth thinking in which the poster engaged. Typically, user communities filter posts by their recency, contributor, or views, to help users manage the vast amount of information available (Ma and Agarwal 2007). Our current findings supply community managers and members with an impetus to seek posts that offer high quality argument development. Such a restructuring of posts can help members and managers sift through the informational clutter; according to structural priming research (Bock and Griffin 2000), it also may stimulate other members to contribute similarly. Furthermore, current community member status and recognition systems, which tend to be based on members' post quantity, could be complemented by assessments and rewards of members' well-developed and substantiated arguments.

References

Algesheimer, R., Dholakia, U. M., and Herrmann, A. 2005. "The Social Influence of Brand Community: Evidence from European Car Clubs," *Journal of Marketing* (69:3), pp. 19-34.

Berger, P., and Luckmann, T. 1966. *The Social Construction of Reality*, London: Penguin.

Bock, K., and Griffin, Z. M. 2000. "The Persistence of Structural Priming: Transient Activation or Implicit Learning?," *Journal of Experimental Psychology: General* (129:2), pp. 177-192.

Brennan, S. E., and Clark, H. H. 1996. "Conceptual Pacts and Lexical Choice in Conversation," *Journal of Experimental Psychology: Learning, Memory, and Cognition* (22:6), pp. 1482-1493.

Brockman, B. K., and Morgan, R. M. 2006. "The Moderating Effect of Organizational Cohesiveness in Knowledge Use and New Product Development," *Journal of the Academy of Marketing Science* (34:3), pp. 295-307.

Cacioppo, J. T., and Petty, R. E. 1981. "Social Psychological Procedures for Cognitive Response Assessment: The Thought-Listing Technique," in *Cognitive Assessment,* X. V. Merluzzi, C. R. Glass and M. Genest (eds.), New York: Guilford Press, pp. 309-342.

Chidambaram, L. 1996. "Relational Development in Computer-Supported Groups," *MIS Quarterly* (20:2), pp. 143-165.

Chung, C. K., and Pennebaker, J. W. 2007. "The Psychological Function of Function Words," in *Social Communication: Frontiers of Social Psychology,* K. Fiedler (ed.), New York: Psychology Press, pp. 343-359.

Cohn, M. A., Mehl, M. R., and Pennebaker, J. W. 2004. "Linguistic Indicators of Psychological Change after September 11, 2001," *Psychological Science* (15), pp. 687-693.

Cole, M. S., Bedeian, A. G., Hirschfeld, R. R., and Vogel, B. 2011. "Dispersion-Composition Models in Multilevel Research," *Organizational Research Methods* (14:4), pp. 718-734.

Dahlander, L., and Frederiksen, L. 2012. "The Core and Cosmopolitans: A Relational View of Innovation in User Communities," *Organization Science* (23:4), pp. 988-1007.

DeKinder, J. S., and Kohli, A. K. 2008. "Flow Signals: How Patterns over Time Affect the Acceptance of Start-Up Firms," *Journal of Marketing* (72:5), pp. 84-97.

Dennis, A. R., Fuller, R. M., and Valacich, J. S. 2008. "Media, Tasks, and Communication Processes: A Theory of Media Synchronicity," *MIS Quarterly* (32:3), pp. 575-600.

DeSanctis, G., and Monge, P. 1999. "Communication Processes for Virtual Organizations," *Organization Science* (10:6), pp. 693-703.

Dholakia, U., Bagozzi, R., and Pearo, L. 2004. "A Social Influence Model of Consumer Participation in Network- and Small-Group-Based Virtual Communities," *International Journal of Research in Marketing* (21:3), 09, pp. 241-263.

Di Gangi, P. M., Wasko, M., and Hooker, R. 2010. "Getting Customers' Ideas to Work for You: Learning from Dell How to Succeed with Online User Innovation Communities," *MIS Quarterly Executive* (9:4), pp. 213-228.

Elsbach, K. D., and Bhattacharya, C. 2001. "Defining Who You are by What You're Not: Organizational Disidentification and the National Rifle Association," *Organization Science* (12:4), pp. 393-413.

Fairclough, N. 1992. "Discourse and Text: Linguistic and Intertextual Analysis Within Discourse Analysis," *Discourse and Society* (3), pp. 193–217.

Faraj, S., Jarvenpaa, S. L., and Majchrzak, A. 2011. "Knowledge Collaboration in Online Communities," *Organization Science* (22:5), pp. 1224-1239.

Farzan, R., Kraut, R., Pal, A., and Konstan, J. 2012. "Socializing Volunteers in an Online Community: A Field Experiment," in *Proceedings of the 2012 Conference on Computer Supported Cooperative Work*, New York: ACM Press, pp. 1-10.

Fayard, A.-L., and DeSanctis, G. 2010. "Enacting Language Games: The Development of a Sense of 'We-ness' in Online Forums," *Information Systems Journal* (20:4), pp. 383-416.

Giles, H. 2009. "The Process of Communication Accomodation," in *The New Reader in Sociolinguistics,* N. Coupland and A. Jaworski (eds.), Basingstoke, England: Macmillan, pp. 276-286.

Giles, H., Willemyns, M., Gallois, C., and Anderson, M. C. 2007. "Accommodating a New Frontier: The Context of Law Enforcement," in *Social Communication,* K. Fiedler (ed.), Hove, England: Psychology Press, pp. 129-162.

Gill, A. M., and Whedbee, K. 1997. "Rhetoric," in *Discourse Studies: A Multidisciplinary Introduction,* T. A. van Dijk (ed.), Beverly Hills, CA: Sage, pp. 157-183.

Gonzales, A. L., Hancock, J. T., and Pennebaker, J. W. 2010. "Language Style Matching as a Predictor of Social Dynamics in Small Groups," *Communication Research* (37:1), pp. 3-19.

Gouran, D. S. 1990. "Exploiting the Predictive Potential of Structuration Theory," in *Communication Yearbook,* J. A. Anderson (ed.), Newbury Park, CA: Sage, pp. 313-322.

Gumperz, J. J. And Levinson, S. C. (eds.). 1996. *Rethinking Linguistic Relativity*. Cambridge, UK: Cambridge University Press.

Hansen, M. T., and Haas, M. R. 2001. "Competing for Attention in Knowledge Markets: Electronic Document Dissemination in a Management Consulting Company," *Administrative Science Quarterly* (46:1), pp. 1-28.

Hardy, C., Lawrence, T. B., and Grant, D. 2005. "Discourse and Collaboration: The Role of Conversations and Collective Identity," *The Academy of Management Review* (30:1), pp. 58-77.

Harrison, D. A., and Klein, K. J. 2007. "What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations," *The Academy of Management Review* (32:4), pp. 1199-1228.

Heracleous, L., and Marshak, R. J. 2004. "Conceptualizing Organizational Discourse as Situated Symbolic Action," *Human Relations* (57:10), pp. 1285-1312.

Herring, S. C. 2001. "Computer-Mediated Discourse," in *The Handbook of Discourse Analysis,* D. Tannen, D. Schiffrin, and H. Hamilton (eds.), Oxford, UK: Blackwell, pp. 612-635.

Huffaker, D. A., Swaab, R., and Diermeier, D. 2011. "The Language of Coalition Formation in Online Multiparty Negotiations," *Journal of Language and Social Psychology* (30:1), pp. 66-81.

Ireland, M. E., and Pennebaker, J. W. 2010. "Language Style Matching in Writing: Synchrony in Essays, Correspondence, and Poetry," *Journal of Personality and Social Psychology* (99:3), pp. 549-571.

Iriberri, A., and Leroy, G. 2009. "A Life-Cycle Perspective on Online Community Success," *ACM Computing Surveys* (41:2), pp. 11-29.

James, L. R., Demaree, R. G., and Wolf, G. 1984. "Estimating Within-Group Interrater Reliability with and Without Response Bias," *Journal of Applied Psychology* (69:1), pp. 85-98.

Jokisaari, M., and Nurmi, J. E. 2009. "Change in Newcomers' Supervisor Support and Socialization Outcomes after Organizational Entry," *The Academy of Management Review* (52:3), pp. 527-544.

Kammeyer-Mueller, J. D., Wanberg, C. R., Glomb, T. M., and Ahlburg, D. 2005. "The Role of Temporal Shifts in Turnover Processes: It's about Time," *Journal of Applied Psychology* (90:4), pp. 644-658.

Kock, N. 2004. "The Psychobiological Model: Towards a New Theory of Computer-Mediated Communication Based on Darwinian Evolution," *Organization Science* (15:3), pp. 327-348.

Koh, J., Kim, Y. G., Butler, B., and Bock, G. W. 2007. "Encouraging Participation in Virtual Communities," *Communications of the ACM* (50:2), pp. 68-73.

Kohler, T., Fueller, J., Matzler, K., and Stieger, D. 2011. "Co-creation in Virtual Worlds: The Design of the User Experience," *MIS Quarterly* (35:3), pp. 773-788.

Kozlowski, S. W. J., and Ilgen, D. R. 2006. "Enhancing the Effectiveness of Work Groups and Teams," *Psychological Science in the Public Interest* (7:3), pp. 77-124.

Kreiner, G. E., and Ashforth, B. E. 2004. "Evidence Toward an Expanded Model of Organizational Identification," *Journal of Organizational Behavior* (25:1), pp. 1-27.

Levine, J. M., and Moreland, R. L. 1994. "Group Socialization: Theory and Research," *European Review of Social Psychology* (5:1), pp. 305-336.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oaks, CA: Sage Publications.

Ma, M., and Agarwal, R. 2007. "Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities," *Information Systems Research* (18:1), pp. 42-67.

Marsman, J. 2012. Personal interview.
Meyerson, D. E., and Scully, M. A. 1995. "Tempered Radicalism and the Politics of Ambivalence and Change," *Organization Science* (6:5), pp. 585-600.
Moran, E., and Gossieaux, F. 2010. "Marketing in a Hyper-Social World: The Tribalization of Business Study and Characteristics of Successful Online Communities," *Journal of Advertising Research* (50:3), pp. 232-240.
Muniz, A. M., and O'Guinn, T. C. 2001. "Brand Community," *Journal of Consumer Research* (27:4), pp. 412-432.
Niederhoffer, K. G., and Pennebaker, J. W. 2002. "Linguistic Style Matching in Social Interaction," *Journal of Language and Social Psychology* (21:4), pp. 337-360.
Postmes, T., Spears, R., and Lea, M. 2000. "The Formation of Group Norms in Computer-Mediated Communication," *Human Communication Research* (26:3), pp. 341-371.
Ransbotham, S., and Kane, G. C. 2011. "Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia," *MIS Quarterly* (35:3), pp. 613-627.
Raudenbush, S. W., and Bryk, A. S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, Newbury Park, CA: Sage Publications.
Searle, J. 1975. "Indirect Speech Acts," in *Syntax and Semantics 3: Speech Acts,* P. Cole and J. L. Morgan (eds.), New York: Academic Press, pp. 59-82.
Seibold, D. R., Lemus, D. R., and Kang, P. 2010. "Extending the Conversational Argument Coding Scheme in Studies of Argument Quality in Group Deliberations," *Communication Methods and Measures* (4:1-2), pp. 46-64.
Seibold, D. R., and Meyers, R. A. 2007. "Group Argument : A Structuration Perspective and Research Program," *Small Group Research* (38:3), pp. 312-336.
Stonequist, E. 1937. *The Marginal Man*, New York: Russell and Russell.
Straub, D. W. and Ang, S. 2011. "Editor's Comments: Rigor and Relevance in IS Research: Redefining the Debate and a Call for Future Research," *MIS Quarterly*, (35:1), pp. iii-xi.
Tausczik, Y. R., and Pennebaker, J. W. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," *Journal of Language and Social Psychology* (29:1), pp. 24-54.
Wasko, M. M., and Faraj, S. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS Quarterly* (29:1), pp. 35-57.
Weick, K. E. 1995. *Sensemaking in Organizations*, Newbury Park, CA: Sage Publications.

About the Authors

Stephan Ludwig is a senior lecturer at the Department of Marketing and Business Strategy of the University of Westminster, United Kingdom. He received his Ph.D. from Maastricht University in 2012. His research interests are social media and the implications of online conversations for corporate performance, customer co-creation, strategic marketing, and service marketing. He has also worked as a market research consultant for InSites Consulting and YouGov Deutschland. His work is published in international journals (e.g., *Journal of Marketing*) as well as popular research magazines in industry (e.g., *Research-Live, Esomar, Casro)*.

Ko de Ruyter is Professor of Marketing at Maastricht University, the Netherlands. His research interests focus on social media, customer loyalty, and environmental stewardship. He has published six books and numerous scholarly articles in *Journal of Marketing, Management Science, Journal of Consumer Research, Journal of Retailing, Journal of the Academy of Marketing Science, International Journal of Research in Marketing, Decision Sciences, Organization Science, Marketing Letters, Journal of Management Studies, Journal of Business Research, Journal of Economic Psychology, Journal of Service Research, Information and Management, European Journal of Marketing,* and *Accounting, Organisation and Society*.

Dominik Mahr is an assistant professor of strategic marketing and innovation management at the School of Business and Economics, Maastricht University, the Netherlands. His research interests are in open innovation, customer cocreation, digital media, knowledge creation, health care, and strategic marketing. His publications have appeared in journals such as *Journal of Product Innovation Management, Research Policy, Journal of Service Research, BMJ Quality & Safety* (previously *Quality & Safety in Health Care*), *Health Policy*, and *Journal of Service Management*.

Martin Wetzels is Professor of Marketing and Supply Chain Research, and head of Department in the Department of Marketing and Supply Chain Management at Maastricht University, the Netherlands. His main research interests are services marketing, marketing research, B2B marketing, marketing channels, and digital marketing. His work has been published in *MIS Quarterly, Journal of Marketing, Management Science, Journal of Consumer Research, Journal of the Academy of Marketing Science, Journal of Retailing, Journal of Service Research, Journal of Consumer Psychology, Journal of Business Research,* and *International Journal of Research in Marketing.* He has contributed more than 80 papers to conference proceedings.

Elisabeth (Lisa) Brüggen, is an associate professor at the Department of Marketing & SCM at Maastricht University and a research fellow at Netspar, the Network for Studies on Pensions, Aging and Retirement. She received her Ph.D. from Maastricht University in 2006 and has research interests in services marketing and online marketing research. Her work has been published in *Journal of Marketing, Journal of Service Research, Journal of Interactive Marketing*, and *Marketing Letters*, among others.

Tom de Ruyck is head of Consumer Consulting Boards at InSites Consulting and in charge of InSites' global activities in terms of community research on thought leadership, steering innovation, and business development. Tom is a frequent presenter at international marketing, research and technology conferences. He is coauthor of *'The Consumer Consulting Board* and has published over 50 white papers, articles in academic journals, business books, and trade magazines. In addition to his work at InSites Consulting, Tom is a guest lecturer at the Vlerick Business School (Belgium) and TiasNimbas (the Netherlands), and an adjunct professor at IESEG School of Management (France) and EHSAL Management School (Belgium).