A New Metric for Assessing Group Level Participation in Fluid Teams

Susannah B. F. Paletz and Christian D. Schunn
LRDC 816
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260
sbfpaletz@gmail.com

***Abstract*-Equality of participation is an important factor in the success of multidisciplinary science teams. The typical measure, standard deviation, fails to provide unbiased estimates across groups of different sizes or within groups that change size over time. We propose a new metric of participation equality that takes into account real-world teams that have members come and go naturally over the course of a meeting. This new metric ranges from zero (entirely equal participation) to one (entirely dominated by a single person). This metric is at the group level and for whatever period of time the researcher specifies. Using 11 hours of transcribed utterances from informal, fluid, co-located meetings during the Mars Exploration Rover (MER) mission, we computed this metric for 549 blocks of time. We found that this metric had good convergent validity via having strong positive correlations with both a standard deviation metric of words spoken and participation equality as assessed by two independent coders. It also had good discriminant validity by being uncorrelated with positive and negative affect words, including anxiety and sadness words. Furthermore, when only fluid groups were examined, it maintained a strong correlation with coder-assessed participation. Future research can take advantage of this metric in other settings where team membership is fluid and equality of participation is of interest.**

INTRODUCTION

Science and engineering teams are increasingly multidisciplinary. Funding agencies have recognized that solving complex problems often requires teams from multiple disciplines, and universities are continuing to develop cross-disciplinary programs [1]. Although multidisciplinary teams hold a great deal of promise, they also are often fraught with difficulty. By integrating the diverse social and cognitive psychological literatures, we recently elaborated a model of multidisciplinary team innovation [2]. We discuss how a central social variable, participation, is important for multidisciplinary team innovation, and then describe a new, group level metric we created in order to assess equality of participation in fluid teams. Participation is the degree to which members of a group talk and share information. At a time when psychology is only rarely measuring actual behavior [3], it is particularly important to create a logistically simple metric for this important variable.

Participation has been studied by social scientists from a range of disciplines. Communication researchers have typically studied participation as a phenomenon for its own sake. A great deal of interesting work has pointed out that communication is necessarily relational, and researchers have developed methods for coding statements for their interactions and reactions to what the person before said (e.g., [4], [5]). Some studies have examined the tendency for certain team members to dominate the conversation, whereas others have presented and tested computational models of participation and turn taking (e.g., [6]). For example, the longer a conversational partner talked during a speaking turn, the more he or she was perceived as interpersonally dominant (e.g., [7], [8]). Within social and organizational psychology, however, participation, and equality of participation in particular, has been examined as a key factor in team performance and decision making.

A. The Role of Participation in Harnessing Knowledge Diversity for Innovation

Knowledge diversity has been implicated as a positive factor in team innovation, but studies of its effects have failed to find consistent results [9]. Knowledge diversity is thought to be critical to complex performance, but is dependent upon communication methods [10]. Our model of team innovation explained the inconsistent results for knowledge diversity examining the critical role of participation and by separating out divergent and convergent thinking processes [2]. We proposed that the social and cognitive implications are different for each type of thinking process, providing a possible explanation for knowledge diversity's mixed results on team outcomes. A number of variables were mapped out; especially relevant for the current paper is the role of sufficient participation.

Vital for both the divergent and convergent thinking paths is sufficient participation, especially when it involves information sharing. For example, knowledge diversity is thought to be positively associated with team innovation via the team having access to a broader range of perspectives, information, and opinions [9], [11], [12]. Knowledge diversity is further theorized to be particularly important

This research was supported in part by a National Science Foundation Grant #SBE-0830210 to the authors through the Science of Science and Innovation Policy Program. The opinions expressed are those of the authors and not necessarily those of NSF.

when the task requires information processing and innovative solutions [13], such as in science teams. These different perspectives must, however, be communicated across the group. Only via sufficient participation and information sharing, which are important for groups whose members hold unshared information, can the team take advantage of the diversity of background information.

Another type of team structure, formal roles, is useful in encouraging participation insomuch as the formal roles have associated communication norms. One type of formal role in multidisciplinary science teams is the local expert (e.g., the immunologist on a team examining a complex health syndrome). When experts were assigned rather than assumed in an experiment, groups were more likely to discuss unshared information [14]. Thus, communication norms such as respecting expert opinion are useful for taking advantage of multidisciplinary teams' breadth of knowledge.

Research on team performance in general has also suggested the importance of sufficient participation. In a study of four-person groups playing a complex team video game, the most successful groups also had more equal participation than the unsuccessful teams [15].

Participation is also essential to several cognitive factors that are important to innovation. Three key cognitive processes are a search for information, analogy, and evaluation [2]. In studies of group cognition, knowledge diversity in microbiology labs has been associated with an improved ability to generate useful analogies [16], [17], assumedly because team members brought a variety of background knowledge to bear. Evaluation is necessary to discriminate between different ideas and picking the best one: Without a shared vision of the problem constraints, groups cannot come to a collective understanding of what constitutes good quality, and thus cannot choose the most innovative outcome [18]. Information sharing is necessary to create this common understanding.

B. Measuring Participation

Unlike in the communication literature where each utterance is examined for its relational purpose, research examining the role of participation on team performance requires a group-level (and/or block of time-level) metric. In addition, it is not uncommon to measure group level participation as equality rather than equity of communication. Equity—to each according to his/her need, from each according to his/her knowledge/ability—would be the best measure of sufficient information sharing, but it is difficult to measure in real world settings without first thoroughly assessing participants' background knowledge. In experiments, it is possible to measure equity: In the hidden profile research paradigm, researchers can manipulate the information each participant has and then determine whether the proportion of the time different kinds of information is discussed (e.g., [19]).

In settings where information is not experimentally manipulated, researchers often rely on measuring *equality* of communication, specifically by measuring the standard deviation of number of remarks or word count (e.g., [15], [20]). The larger the standard deviation, the more inequality in how much each individual spoke, and the more likely that one person dominated over the others. These measures are useful in that they are easy to calculate, given information about who said how many words or utterances. This measure is predicated, however, on group size being stable. In the Fischer et al. study [15], the teams were composed of four individuals, no more, no less. In many real world settings the number of meeting participants fluctuates. For example, in the Mars Exploration Rover (MER) mission, scientists were co-located for the first 90 Martian days of the project. Some of their meetings were formal and had a stable number of participants within a meeting, but different formal meetings had different numbers of participants. Many meetings were informal, occurring naturally as scientists walked up to others' workstations and started conversations. In these informal groups, two individuals might be sitting at their workstations chatting when a third scientist approaches them; a fourth might join, and then one of the original scientists might then leave. In analyzing transcripts from the MER mission in order to test our model [2], we recognized that a new metric of participation was necessary that took into account the fluid nature of the informal conversations and the variable group size of all meetings.

The fluid nature of meeting membership necessitated, first, that we note how many people were present at any one moment during the conversation. The derived metric we created can be used for any length of transcript. It was also formulated to be used on transcripts where clauses (utterances) were separated into separate lines, much as in cognitive psychology research where each thought is coded. These utterances are potentially shorter than turns, and a person could speak several utterances consecutively.

$$P_s = \bar{n}^2 \bullet \frac{\sum_{i=1}^{N} \left| \sum_{K=1}^{M} f(n_k, i, K) \right|}{2(\bar{n}-1)\sum_{i=1}^{N} m_i} \tag{1}$$

Where:

n_k is the number of people present on utterance (line) K

i is the index of a particular person in the group

M is the maximum number of utterances spoken in the block, segment, or clip being studied

m_i is the number of utterances (lines) when person i is present

N is the number of people ever present in the block, segment, or clip being studied

$\bar{n}$ is the average number of people present per utterance in the block, segment, or clip being studied

and

$$f(n_k, i, K) = \begin{cases} \dfrac{n_k - 1}{n_k}, & \text{if } i \text{ is present, speaking utterance K} \end{cases}$$

$$f(n_k, i, K) = \left\{ \frac{n_k - 1}{n_k}, \right.$$ if i is present and *silent* during utterance K

$$f(n_k, i, K) = \{0,$$ if i is absent during utterance K

The first step in computing this metric is to compute the function for the three conditions detailed above for each utterance or line of text spoken. This aspect of the metric takes into account variable group size, weighting each person's utterance based on how many people are present at that moment. The more people present, the heavier the person speaking is weighted, whereas those not speaking are penalized less. For example, if two people are present, the speaker gets given the number ½, and the other -½. If three are present, the speaker gets 2/3 and each silent party gets -1/3. Similarly, if six people are present, the speaker gets 5/6 for that utterance and each silent person is given -1/6. Individuals not present during that part of the conversation (i.e., people who show up later) are given a zero (see Table 1).

TABLE 1
STEPS ONE THROUGH THREE, ONE EXAMPLE

Utterance	Speaker	Person					
		1	2	3	4	5	6
Have you received the file yet?	1	1/2	-1/2	0	0	0	0
No	2	-1/2	1/2	0	0	0	0
Wait, which file?	2	-1/3	2/3	-1/3	0	0	0
Have you guys seen the file?	3	-1/3	-1/3	2/3	0	0	0
No	2	-1/3	2/3	-1/3	0	0	0
Yes	1	2/3	-1/3	-1/3	0	0	0
I just sent this great document,	4	-1/6	-1/6	-1/6	5/6	-1/6	-1/6
did you guys get it?	4	-1/6	-1/6	-1/6	5/6	-1/6	-1/6
It's about the new picture we just got.	4	-1/6	-1/6	-1/6	5/6	-1/6	-1/6
Step 2, Sum across utterances		-0.83	0.17	-0.83	2.50	-0.50	-0.50
Step 3: Absolute Value of Step 2		0.83	0.17	0.83	2.50	0.50	0.50

In the second step, each person's speaking-or-not numbers given in the first step are summed across all utterances in the block of time analyzed (say, 25 utterances, about one minute, see Table 1). The third step involves taking the absolute value of each speaker's sum across the utterances (see Table 1, last row). The next element is to compute a weighted average of the participants. This takes into account an important issue: Not every individual is present for the entire block of utterances. If one individual shows up for a brief amount of time, makes a request, and leaves, that activity would skew the formula toward dominance, even if the majority of the block involved a fairly equal conversation between other individuals. This is accomplished by steps four and five: The fourth step computes a sum across all the individuals who are present during the block of time with their respective numbers created by the third step. This gives a number that is truly at the group and block of time level. Then, the fifth step involves controlling for the number of lines each person is present. This is done via summing the total of the number of lines each person is present, and dividing the result of step four by this number (1). Finally, in order for the metric to range from zero to one, it needs to control for the maximum amount possible. This is accomplished by the last step, step six, which involves multiplying the number generated by the fifth step by the following formula (2), which is embedded in (1):

$$\frac{\bar{n}^2}{2(\bar{n} - 1)} \tag{2}$$

As noted above, $\bar{n}$ is the average number of people present per utterance in the block, segment, or clip being studied. This is created by taking every utterance and counting how many people are present at that time, and then taking the average of that across the block or segment that is being analyzed. Thus, the new measure, labeled the 'ParticipaSchunn metric', ranges from zero to one, with zero for entirely equal participation, given a variable group size, and one for complete dominance of the discussion block by any single individual. The metric is at the level of the group for a particular block of time. Table 2 shows examples of what occurs in each of the steps four through six, including the example in Table 1. Other examples include mostly equal and strongly, but not completely, dominated groups of two, three, and six, as well as mostly equal and strongly dominated fluid groups of two to four.

TABLE 2
STEPS FOUR THROUGH SIX

Type of Group	Step 4	Step 5	Step 6: ParticipaSCHUNN (0 to 1, equal to dominated)
Mixed, Fluid (Table 1 Example)	5.33	0.16	0.40
Dominated 2-Person Group	8.00	0.33	0.67
Dominated 3-Person Group	12.00	0.33	0.75
Dominated 6-Person Group	16.00	0.22	0.80
Dominated, Fluid 2-4 Person Group	14.50	0.35	0.85
(Almost) Equal 2-Person Group	2.00	0.08	0.17
Equal 3-Person Group	0.00	0.00	0.00
Equal 6-Person Group	0.00	0.00	0.00
(Almost) Equal, Fluid 2-4 Person Group	2.50	0.06	0.15

In this study, we set out to test the discriminant and convergent validity of this metric. We utilized the MER data mentioned above. Convergent validity was determined by measuring equal participation in two other ways, and discriminant validity was ascertained by measuring a variety of affect variables. We chose affect words because they can be easily identified in text. More importantly, we have no reason to believe affect is correlated with equality or dominance of participation. An entire group can be excited or upset and sharing it equally; on the other hand, a single individual could be sharing their feelings with the rest of the group.

METHOD

A. Participants

The overall MER science team had over 100 members during the first 90 Martian days of the mission; the contact list in 2006 listed 238 members. Given the video quality of the clips we had available, it was not consistently possible to distinguish between participants. Audio-video clips were chosen based on whether the conversation could be heard and whether it was related to the MER mission. Clips began and ended based on audibleness of the speakers but also whether a conversation naturally began or was completed. Conversations often ended naturally by the speakers re-focusing on their desktop computers or leaving the area. The 114 clips we used ranged from two to ten participants. Because the conversations during the clips flowed and changed both in terms of topic content and number of conversationalists, we analyzed the data at the level of the segmented block (see below for a description), rather than the clip.

The transcripts were put into Excel where each line was an utterance (thought statement, see Table 1 for a created example). In total, we coded 12,336 utterances/lines of transcript, or roughly 11 hours and 25 minutes of conversation. Although clips were chosen for being on-topic, we also coded whether they were substantive talk or not (kappa = .96). Substantive talk included discussions of process and relationships issues relevant to the MER mission, but did not include conversations irrelevant to MER (e.g., a brief discussion of iPods). The analyses were conducted on the remaining on-topic talk, which comprised 11,856 utterances and about 11 hours of conversation.

B. Measures

Participation was ascertained in three ways: (1) through having two independent coders rate the participation on a 0 (completely equal participation) to 100 (completely dominated by one person) scale, (2) through taking the standard deviation (SD) of number of words spoken per person based on the maximum number of people present in a block of text, divided by the total number of words in that block to control for block size, and (3) via the new metric (ParticipaSchunn) described above that took into account the constantly changing number of participations to have total equal participation (0) and totally dominated by one person (1). The alpha reliability for the coded participation was .95, with a single measures two-way mixed model intraclass correlation (ICC) of .90 (95% confidence interval from .89 to .92).

To measure affect, we utilized Pennebaker, Booth, and Francis's Linguistic Inquiry and Word Count (LIWC, [21]), a computer program that identifies specific affect words in text [22]. These include both positive and negative affect words. Negative affect word sets include anxiety (e.g., worried, nervous), anger (e.g., annoyed, hate), and sadness (e.g., grief, sad, crying). The LIWC has been used successfully in other team studies that include the analysis of conversational transcripts (e.g., [15]). For these analyses, the LIWC counts for positive and negative affect, anxiety, and sadness were divided by the total number of words per block (as estimated by the LIWC) before being correlated with the ParticipaSCHUNN metric. Anger words occurred too rarely to be used in these analyses (mean of anger words by total number of words = 0).

C. Analyses

Each utterance was on a different line, resulting in 11,856 lines. We also sought a level of analysis between the level of the clip, which we felt could contain too many topics, and the individual line/utterance, which could not be used to measure participation. Clips were broken up into *blocks* of utterances based on whether they included analogies for a separate study. Blocks were no more than 25 utterances long, or about a minute in time. Blocks of fewer than 5 utterances were removed from the analyses, resulting in 549 different blocks for each analysis. Because all of the variables were non-normal (Shapiro-Wilks statistic significant at < .001), Spearman Rho's correlations were used.

RESULTS

First, we examined whether the three different participation measures correlated together. The correlations between the ParticipaSchunn metric and the coded participation and standard deviation-derived measures were significant and positive (see Table 3). These correlations suggest that a modicum of convergent validity for the ParticipaSCHUNN metric.

Although many of the affect variables were correlated with each other, the ParticipaSCHUNN metric was not significantly correlated with positive affect, negative affect, or two of negative affect's specific components—anxiety or sadness (see Table 3).

In addition, we examined the correlations between the three participation metrics under two conditions: when the groups were fluid and when they were stable, i.e., there was no change in group membership during the block. When blocks of time involved groups that were stable, all three correlated highly together (see Fig. 1: n = 386, ParticipaSchunn & coded participation, r_s = .81, p < .001; ParticipaSchunn and standard deviation metric, r_s = .75, p < .001; and coded participation and standard deviation metric, r_s = .80, p < .001). When groups were fluid such that members came and went during the course of the conversation, participaSCHUNN was still highly correlated with participation as judged by coders, but participaSCHUNN and standard deviation of words were not correlated as strongly as before. Coded participation was not as highly correlated with standard deviation of words as it was with participaSCHUNN (see Fig. 1: n = 163, ParticipaSchunn & coded participation, $r_s = .64$, $p < .001$; ParticipaSchunn and standard deviation metric, $r_s = .34$, $p < .001$; and coded participation and standard deviation metric, $r_s = .54$, $p < .001$). Thus, the value of using the new metric is clear in the variable membership case, the case for which it was designed.

TABLE 3
CORRELATIONS, MEANS, AND STANDARD DEVIATIONS OF PARTICIPATION METRICS AND AFFECT WORDS

Variable	*M* (*SD*)	Correlations (Spearman Rho) 1	2	3	4	5	6
1. Participa-SCHUNN	.40 (.21)						
2. Standard deviation of words	.26 (.13)	.66**					
3. Coded participa-tion	52.19 (21.39)	.76**	.74**				
4. Positive affect words	.02 (.04)	-.06	< .01	-.07			
5. Negative affect words	.01 (.02)	.02	.03	.05	-.05		
6. Anxiety words	< .01 (< .01)	.02	-.02	.03	-.08	.37**	
7. Sadness words	< .01 (<.01)	-.01	-.05	-.02	-.09*	.40**	.05

* p < .05, ** p < .001, N = 549

DISCUSSION

We successfully created a new metric that takes into account the fluid membership of natural group conversations that occur in some real world settings. It was strongly positively correlated with other measures of equality of participation—standard deviation of words spoken and coder judged equality of participation—and not correlated with affect, as expected. Further research in this area should test the use of this metric in other settings with fluid membership of team conversations. In teams at work, this metric could be used to study other co-located groups that have fluid, informal meetings, such as agile software development teams, other space missions, and water cooler conversations. Outside of work, this metric could be utilized in studying conversations of people at leisure such as public places, in parks, or at private parties.

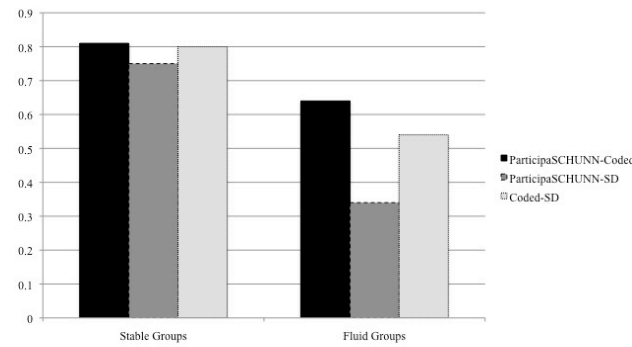

Fig. 1. Correlations between ParticipaSCHUNN, coded participation, and standard deviation of words (SD) by stable versus fluid groups.

In laboratory studies, the experimenter controls the amount of information held by each participant such that equal participation is often artificially created to be ideal. In natural settings, the distribution of useful information is unknown, changing, and/or unequal, suggesting that sufficient information may not be pure equality. Instead, equity may be optimal. This metric can also help measure *degrees* of equality and dominance. For example, if a researcher knows that group members hold unequally important background information, different levels of equality and dominance of participation can be hypothesized to be important. A researcher could use this metric to test hypotheses involving ideal but alternative types of participation levels.

Equal participation and information sharing are fundamental social psychological processes in teams. Even more importantly, they are necessary for multidisciplinary teams to take advantage of their background knowledge in order to be innovative. With this new metric, we can now measure participation in natural, fluid group conversations. The applications of this metric go beyond our study of team innovation to any setting where researchers wish to measure participation in changing groups.

ACKNOWLEDGMENT

The authors are grateful to K. Glassman, J. Krill, and T. Wong for their assistance on this project.

REFERENCES

[1] S. J. Derry, C. D. Schunn, C. D., and M. A. Gernsbacher, Eds. *Interdisciplinary Collaboration: An Emerging Cognitive Science*. Mahwah, NJ: Erlbaum, 2005.

[2] S. B. F. Paletz and C. D. Schunn, "A social-cognitive framework of multidisciplinary team innovation," *Topics in Cognitive Science*, In press.

[3] R. F. Baumeister, K. D. Vohs, and D. C. Funder, "Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?," *Perspectives on Psychological Science,* vol. 2, pp. 396-403, 2007.

[4] F. E. Millar and L. E. Rogers, “A relational approach to interpersonal communication,” in *Explorations in Interpersonal Communication*, G. R. Miller, Ed. Beverly Hills, CA: Sage, 1976, pp. 87-103.

[5] L. E. Rogers and R. V. Farace, “Analysis of relational communication in dyads: New measurement procedures,” *Human Communication Research,* vol. 1, pp. 222-239, 1975.

[6] G. Stasser and S. I. Vaughan, “Models of participation during face-to-face unstructured discussion,” in *Understanding Group Behavior: Consensual Action by Small Groups*, Vol. 1, E. H. Witte and J. H. Davis, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, pp. 165-192.

[7] J. P. Folger, “The effects of vocal participation and questioning behavior on perceptions of dominance,” *Social Behavior and Personality,* vol. 8, pp. 203-207, 1980.

[8] M. T. Palmer, “Controlling conversations: Turns, topics, and interpersonal control,” *Communication Monographs,* vol. 56, pp. 1-18, 1989.

[9] E. Mannix and M. A. Neale, “What differences make a difference? The promise and reality of diverse teams in organizations,” *Psychological Science in the Public Interest*, vol. 6, no. 2, pp. 31-55, 2005.

[10] T. Okada and H. A. Simon, “Collaborative discovery in a scientific domain,” *Cognitive Science,* vol. 21, pp. 109–146, 1997.

[11] B. A. Nijstad and W. Stroebe, “How the group affects the mind: A cognitive model of idea generation in groups,” *Personality and Social Psychology Review,* vol. 10, pp. 186-213, 2006.

[12] D. van Knippenberg and M. C. Schippers, “Work group diversity. *Annual Review of Psychology*, vol. 58, pp. 515-541, 2007.

[13] D. van Knippenberg, C. K. W. De Dreu, and A. C. Homan, “Work group diversity and group performance: An integrated model and research agenda,” *Journal of Applied Psychology,* vol. 89, pp. 1008-1022, 2004.

[14] D. D. Stewart and G. Stasser, “Expert role assignment and information sampling during collective recall and decision making,” *Journal of Personality and Social Psychology,* vol. 69, pp. 619-628, 1995.

[15] U. Fischer, L. McDonnell, and J. Orasanu, “Linguistic correlates of team performance: Toward a tool for monitoring team functions for space missions,” *Aviation, Space, and Environmental Medicine*, vol. 78, pp. B86-B95, 2007.

[16] K. Dunbar, “How scientists really reason: Scientific reasoning in real-world laboratories,” in *The Nature of Insight,* R. J. Sternberg and J. E. Davidson, Eds. Cambridge, MA: The MIT Press, 1995, pp. 365-395.

[17] K. Dunbar, “How scientists think: On-line creativity and conceptual change in science,” in *Creative Thought: An Investigation of Conceptual Structures and Processes,* T. B. Ward, S. M. Smith, and J. Vaid, Eds. Washington: American Psychological Association, 1997, pp. 461-493.

[18] A. Cropley, “In praise of convergent thinking,” *Creativity Research Journal,* vol. 18, pp. 391-404, 2006.

[19] S. Schulz-Hardt, F. C. Brodbeck, A. Mojzisch, R. Kerschreiter, and D. Frey, “Group decision making in hidden profile situations: Dissent as a facilitator for decision quality,” *Journal of Personality and Social Psychology,* vol. 91, pp. 1080-1093, 2006.

[20] J. Siegel, V. Dubrovsky, S. Kiesler, and T. McGuire, ”Group processes in computer-mediated communication,” *Organizational Behavior and Human Decision Processes,* vol. 37, pp. 157-187, 1986.

[21] J. W. Pennebaker, R. J. Booth, and M. E. Francis. LIWC2007. [http://www.liwc.net/index.php]. Austin, TX: University of Texas, 2007.

[22] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, “Psychological aspects of natural language use: Our words, our selves,” *Annual Review of Psychology,* vol. 54, pp. 547-577, 2003.