Title
Distributed Reasoning: An Analysis of Where Social and Cognitive Worlds Fuse

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 18(0)

Authors
Dama, Mike
Dunbar, Kevin

Publication Date
1996

Peer reviewed

Distributed Reasoning: An Analysis of Where Social and Cognitive Worlds Fuse

Mike Dama and **Kevin Dunbar**
Department of Psychology
McGill University
1205 Dr. Penfield Ave. West
Montreal, Quebec H3A 1B1 Canada
dama@ego.psych.mcgill.ca or dunbar@ego.psych.mcgill.ca

Abstract

The goal of this paper was to examine the influence of social and cognitive factors on distributed reasoning within the context of scientific laboratory meetings. We investigated whether a social factor, status, and cognitive factors such as discussion topic and time orientation of the research influenced distributed reasoning. The impact of status on distributed reasoning was examined using 3 lab meetings in which a technician presented (low status) and 3 lab meetings in which a graduate student presented (high status). Two cognitive variables were also examined; focus of discussion topic (theory, method, findings, and conclusions) and the time orientation of the distributed reasoning (past, current and future research). Pooled (cross sectional/time series) analysis, a regression technique, was used to perform the analyses. We found that status of the presenter influenced the structure of distributed reasoning: When the presenter was of high status, the principal investigator was an important influence on distributed reasoning. In contrast, when the presenter was of low status, other lab members were more likely to contribute to distributed reasoning. Our analyses also show that distributed reasoning is not influenced by the discussion topic but appears to focus on the discussion of future research.

Distributed Reasoning:

An Analysis of Where Social and Cognitive Worlds Fuse

Contemporary models of human reasoning have focused on the reasoning processes that individual subjects engage in while they perform a task in a psychological experiment (e.g., Anderson, 1993; Newell & Simon, 1972). While these models have been very successful in accounting for subject's performance in psychology experiments, a number of researchers have turned to understanding the types of reasoning strategies that people use in real-world situations (e.g., Dunbar, 1995, 1996; Hutchins, 1990; Olson et al. 1992). All of these researchers have found that a large amount of reasoning in real-world contexts occurs in groups rather than in individuals. The finding that reasoning occurs in groups raises a number of important questions concerning the nature of group reasoning. One question that we have been investigating is the way components of reasoning are shared among members of a group. We have found that reasoning in groups is distributed over individuals rather than residing within one individual's head (Dunbar, 1995, 1996). That is, different components of a reasoning process, such as an induction, can be performed by a number of individuals. Up until now, we have not investigated the constraints on cognition imposed by social factors in distributed reasoning. The goal of the research reported in this paper is investigate the influence of social and cognitive factors on distributed reasoning.

For the past 5 years we have been investigating the reasoning processes that scientists use in their day-to-day research and have found that distributed reasoning plays a crucial role (Dunbar 1995, 1996; Dunbar, Patel, Baker & Dama, 1995). One place where distributed reasoning is particularly important is in the weekly laboratory meetings that scientists have. We have found that distributed reasoning helps scientists quickly solve problems, propose hypotheses, and design experiments. Analyses of distributed reasoning episodes reveals that one of the main cognitive events that occurs during distributed reasoning is the parallel generation of multiple representations. Having multiple representations available makes it possible for a scientist to quickly search for a representation that will solve their current problem. In this presentation, we will turn to the social mechanisms involved in distributed reasoning. Much research on groups in social psychology has shown that the status of members in a group, and the composition of a group influences group acceptance and discussion of ideas (e.g., Shaw 1981; Weisband, Schneider, & Connolly, 1995). However, this type of research has not addressed the issue of distributed reasoning. Furthermore, given that distributed reasoning is the place that the cognitive and social worlds fuse, it is important to discover the impact of social variables on distributed reasoning. What we will do is analyze laboratory meetings to discover whether social factors such as status have an effect on distributed reasoning and also what the cognitive components of distributed reasoning are. We will also introduce a new statistical tool

Graduate Student		Technician	
Meeting	N of 2 Minute segments	Meeting	N of 2 Minute segments
1	17	1	12
2	22	2	12
3	28	3	38
Totals	67		62

Table 1: Number of observations for each meeting

into the analysis of distributed reasoning that makes it possible to investigate these issues.

This paper will be divided into four main sections. First we will give background information on the meetings that we analyze. Second, we will provide an overview of the statistical technique that we use to analyze distributed reasoning. Third we will present our analyses of distributed reasoning. Finally we will address the issue of how social and cognitive processes interact to produce distributed reasoning.

Distributed Reasoning in Scientific Laboratories

The analysis of distributed reasoning is a component of a research program that is concerned with understanding the cognitive and social processes underlying science as it is practiced. Over the past five years we have collected data from six biological laboratories at two renowned universities, four in the United States, and two in Canada (see Dunbar 1995, 1996 for overviews of the research program). The actual data that we have been analyzing consists of the weekly laboratory meetings that occur in all the labs. We use lab meetings as our data source as the lab meetings are one of the main sources of problem solving, reasoning, and hypothesis generation in the laboratories. For the purpose of this paper we will focus on distributed reasoning in three molecular biology laboratories. All three laboratories are of similar size and use similar techniques. One laboratory is concerned with understanding how a gene in the HIV virus functions, another lab with understanding how a gene works on a parasite, and the third laboratory investigates the function of a particular gene in causing a certain type of cancer. All three of the labs use similar research methods and questions regarding the functions of their respective genes.

Method

Data

The data consisted of 6 audiotaped laboratory meetings. Three of the meetings were where the presenter was of low status (a technician), and three were where the presenter was of higher status (a graduate student). By having presenters of different status, it is thus possible to investigate the role of presenter status in distributed reasoning. In each meeting, current research was presented by either a graduate student or technician. Each audiotaped meeting was transcribed and divided into consecutive 2 minute time segments. For each two minute segment, the number of statements (regardless of whether the statement was coded as one of the variables) were counted. Additionally, within each 2 minute segment, the number of statements were counted for each variable. Each of the variables was converted to a percentage by dividing its count for a time segment by the total number of statements for the time segment. For example, if in a particular time segment 3 distributed reasoning statements occurred out of 10 total statements for the time segment, 30 percent of the time segment was scored as distributed reasoning. Table 1 shows the number of 2 minute segments for each of the six meetings.

Coding of Variables

The variables of interest were generated by coding each statement from the transcribed meetings. Statements consisted of sentences and simple yes/no comments by laboratory members. Each statement was coded for the presence of the following.

(I) Distributed Reasoning. A statement in which a person added a new element to the reasoning of another speaker was coded as distributed reasoning. In addition, any one reasoning episode that was part of a distributed reasoning block was coded as distributed reasoning.

(II) Participation of Principal Investigator/ Others. Statements made by the principal investigator were coded as participation of principal investigator. Statements made by all other participants (excluding the presenter) were coded as participation of others.

(III) Discussion Topic. Four discussion topics were coded. Statements involving background information and current understanding were coded as theory. A statement focusing on how an experiment was conducted was coded as method. Statements involving experimental results were coded as findings. Statements in which the speaker gave an opinion about an experimental result were coded as conclusions.

(IV) Time Orientation of Research. Statements focusing on the discussion of research performed in the past were coded as past research. Statements focusing on the discussion of research in progress or recently completed by the presenter were coded as current research. Statements in which the speaker proposed a methodology, result or a conclusion for a study not yet underway were coded as future research.

Status	Presenter	R Square	B	Beta
PI (high)	Grad Student	.31**	1.14	.56**
	Technician	.16*	.65	.42**
Others (lower)	Grad Student	.21**	.93	.50**
	Technician	.50**	.67	.72**

* $p < .05$. ** $p < .01$.

Table 2: Distributed Reasoning
Predicted From High and Lower Status Lab Members

The PCSTS Statistical Technique

Pooled (cross sectional/time series) analysis was used to analyze distributed reasoning. A parametric regression technique used in econometrics and political science (e.g., Stimson, 1985; Kmenta, 1986), pooled (or PCSTS) analysis was developed to examine data that consists of measurement of variables over time (i.e., a time series) for a number of cross sections or units (e.g., lab meetings). The rationale behind PCSTS analysis is that the researcher may not have enough time series data to warrant the use of most time series models. Additionally, the researcher may not have sufficient cross section data to analyze differences between cross sections. For example, assume we have a data set consisting of 10 time measurements for each of 5 units. Ten time measurements are not enough to perform a time series. Additionally, 5 units are not enough to compare units. In comparison, since PCSTS analysis uses both the time measurements and cross section information, a PCSTS analysis would have 50 observations for analysis, (10 time measurements for each of 5 units). Thus, a PCSTS analysis of time series and cross sectional data makes it possible to increase the number of observations in the analyses. This technique therefore allows us to perform statistical analyses on small data sets such as the data that we analyze in this paper.

Results

PCSTS analysis was performed separately on the graduate student and technician data sets to examine if there were differences in the factors predicting distributed reasoning for graduate students and technicians. To address this question, a regression equation with dummy variables was used. In the dummy variables model, N - 1 dummy variables are entered for the unit of interest, to control for dependent variable differences between units (Stimson, 1985). Thus for each regression equation, 2 dummy variables were entered to control for distributed reasoning differences between the 3 meetings. For a particular equation, the calculated percentages for the variables at each time segment were entered as separate observations in the regression equation. For example, in predicting distributed reasoning from principal investigator comments, the 67 distributed reasoning percentages and their corresponding principal investigator percentages were treated as separate observations for the graduate student meetings.

Question I: Does the status of the presenter influence the structure of distributed reasoning?

The purpose of our first set of analyses was to determine whether the structure of distributed reasoning was influenced by the status of the presenter. Specifically, we were interested in whether the status of the presenter interacted with the contribution by high and low status lab members during distributed reasoning. For a high status presenter, (i.e., a graduate student) two regression equations were run. In the first equation, distributed reasoning was predicted from statements by a high status person (i.e., principal investigator), while in the second equation, distributed reasoning was predicted from statements by lower status lab members (i.e., all other lab members). In addition, the same regression equations were run for a low status presenter (i.e., a technician). The R squares and unstandardized (B) and standardized (Beta) regression weights for the regression equations are given in Table 2.

Examination of the beta weights indicates that statements made by both the principal investigator and others are positively related to distributed reasoning for the graduate student and technician presentations. However, examination of the R squares demonstrates differences between the graduate student and technician presentations. Specifically, statements made by the principal investigator account for more distributed reasoning in the graduate student presentations than in the technician presentations. In contrast, statements made by others account for more distributed reasoning in the technician presentations than in the graduate student presentations. Overall these analyses suggest that the status of the presenter does influence the structure of distributed reasoning.

Question II: Are some topics more likely to be discussed during distributed reasoning than others?

We were interested in determining which of 4 topics (i.e., theory, method, findings, and conclusions) were discussed during distributed reasoning. This question was analyzed by predicting distributed reasoning from discussion of the 4 topics (i.e., theory, method, findings, and conclusions) using separate regression equations for each of the topics.

As can be seen from Table 3, the R squares suggest that none of the 4 discussion topics are strong predictors of distributed reasoning; only one regression equation has an R square above .20. While examination of the significant beta weights indicates that discussion of findings are negatively related to distributed reasoning and discussion of

Topic	Presenter	R Square	B	Beta
Theory	Grad Student	.03	.70	.17
	Technician	.06	-.54	-.19
Method	Grad Student	.01	.10	.09
	Technician	.06	.18	.20
Findings	Grad Student	.10	-.34	-.32**
	Technician	.13*	-.33	-.34**
Conclusions	Grad Student	.12*	.75	.37**
	Technician	.21**	.81	.44**

* $p < .05$. ** $p < .01$.

Table 3: Distributed Reasoning Predicted From Discussion Topics

conclusions are positively related to distributed reasoning, the R squares are fairly low. The reason for the negative relationship between discussion of findings and distributed reasoning is that a number of time segments had high levels of discussion of findings in the absence of distributed reasoning. This resulted in a negative relationship between discussion of findings and distributed reasoning.

Question III: Is distributed reasoning related to the discussion of past, current or future research?

We were interested in determining whether the discussion of past, current or future research was related to distributed reasoning. This question was analyzed by predicting distributed reasoning from discussion of past research, current research and future research, using separate regression equations for each of the 3 time orientations (see Table 4 for the results).

Examination of the R squares indicates that discussion of future research accounts for more distributed reasoning than discussion of current research. An inspection of the beta weights indicates that discussion of current research is negatively related to distributed reasoning, while discussion of future research is positively related to distributed reasoning. The reason for the negative relationship between discussion of current research and distributed reasoning is that a number of time segments had high levels of discussion of current research in the absence of distributed reasoning. This resulted in a negative relationship between discussion of current research and distributed reasoning.

Conclusion

The goal of the analyses performed in this paper was to investigate the interaction of cognitive and social processes in distributed reasoning. We have found that social status does indeed have an effect on when and how distributed reasoning occurs. In contrast, in the technician presentations, the principal investigator making comments does not lead to more distributed reasoning. The presenter status thus influences the flow and structure of distributed reasoning. Currently we are investigating factors which may be responsible for the structure of distributed reasoning. For example, the complexity of the experiments performed by the presenter and presenter expertise may influence the structure of distributed reasoning.

A second finding arising out of our analyses is that that none of the 4 topics (theory, method, findings, and conclusions) were strongly related to distributed reasoning. This result was surprising in that we expected more distributed reasoning to occur when the scientists were discussing the findings of experiments. We are now conducting more detailed analyses where we are examining whether more distributed reasoning occurs during discussion of unexpected findings as compared to expected findings.

Finally, our analysis demonstrated that an important component of distributed reasoning is future research rather than past or current research. Again, this result was surprising. We expected that much of distributed reasoning would focus on current research rather than on research not yet conducted.

On the use of Pooled Cross-Sectional Time series Analysis (PCSTS)

We have introduced PCSTS analysis as a statistical tool in our examination of the laboratory meetings. In particular, we have used an Ordinary Least Squares model with dummy variables to control for distributed reasoning differences between lab meetings. Use of PCSTS analysis allowed us to examine relationships among variables within a limited number of lab meetings.

Time	Presenter	R Square	B	Beta
Past	Grad Student	.01	-.15	-.13
	Technician	.03	-.07	-.06
Current	Grad Student	.18**	-.43	-.51**
	Technician	.32**	-.44	-.58**
Future	Grad Student	.34**	.74	.59**
	Technician	.56**	.71	.73**

* $p < .05$. ** $p < .01$.

Table 4: Distributed Reasoning Predicted From Discussion of Past, Current and Future Research

However, PCSTS has the potential to be used to address many different types of questions that are of interest to cognitive scientists. For example, some PCSTS models allow for the use of variables from earlier time intervals (i.e., lagged predictors) and current time intervals, giving the researcher a method of examining how preceding events influence current events (Sayrs, 1989). We plan to use such models to examine how preceding laboratory meeting events impact later occurrences of distributed reasoning.

Acknowledgments

This research was supported by a McGill Major Graduate Fellowship to the first author and the following grants to The second author was funded by grants number OGP0037356 from NSREC and grants from FCAR and SSHRC. We would like to thank Craig Leth-Steensen for his comments on an earlier version of this paper and his assistance on the use of the PCSTS technique. We would also like to thank Rhonda Amsel for her comments on the manuscript and Romana Ahmad and Annie Maude St-Laurent for their assistance in coding the data.

References

Anderson, J.R. (1993). *Rules of the Mind.* Hillsdale, NJ: LEA.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *Mechanisms of insight.* Cambridge, MA: MIT Press.

Dunbar, K. (1996). How scientists think: Online creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Vaid, *Conceptual structures and processes: Emergence, discovery, and change.* Washington, DC: American Psychological Association Press.

Dunbar, K., Patel, V., Baker, L., & Dama, M. (1995, November). *Group reasoning strategies in knowledge-rich domains.* 36th Annual Meeting of the Psychonomic Society, Los Angeles.

Hutchins, E. (1990). The technology of team navigation. In J. Galegher, R.E. Kraut, & C. Edgido (Ed.), *Intellectual teamwork: Social and technological foundations of cooperative work.* Hillsdale, NJ: LEA.

Kmenta, J. (1986). *Elements of Econometrics* (2nd ed.). New York, NY: Macmillan.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Olson, G.M., Olson, J.S., Carter, M., & Storrsten, M. (1992). Small group design meetings: An analysis of collaboration. *Human Computer Interaction*, 7, 347-374.

Sayrs, L. W. (1989). *Pooled Time Series Analysis.* Beverly Hills, CA: Sage.

Shaw, M. E. (1981). Group Dynamics: The psychology of small group behavior, (3rd ed.). New York, NY: McGraw-Hill.

Stimson, J. (1985). Regression in space and time: A statistical essay. *American Journal of Political Science,* 29, 914-947.

Weisband, S. P, Schneider, S. K, & Connolly, T. (1995). Computer-mediated communication and social information: Status salience and status differences. *Academy of Management Journal* , 38, 1124-1151.