Modelling Policies for Collaboration

Mark Burton (mburton@arm.com)
ARM; 110 Fulbourn Rd
Cambridge CB1 9NJ, UK

Paul Brna (paul@cbl.leeds.ac.uk)
Computer Based Learning Unit; Leeds University
Leeds LS2 9JT, UK

Abstract

There are different ways in which learners can be organised when working together collaboratively. The issue discussed here is how to organise collaborative activities to help students *learn* how to collaborate — in particular, which *policy* is the best one to adopt for managing group communication to achieve this end. Here, different policies are considered which might apply during collaborative task performance. A computation model (Clarissa) is used to examine how these different policies perform in a collaborative task performance. The behaviour of the model is examined from the point of view that participants need to experience a range of dialogue roles which reflect the linguistic functions of utterances that are involved in collaboration. The results derived from the analysis of the behaviour of the computation model complements empirical work on small group behaviour.

Models of Collaboration

Collaboration may be mainly "in the eye of the beholder": in many educational settings the children collaborating are scarcely aware of any rules governing their work together. This leads to the notion that there are degrees of collaboration, that there may be a value for methods of assessing *how much* collaboration there is.

When the process of collaboration is examined, there is an issue about whether the current task is divided into parts tackled by different collaborators or whether collaboration entails synchronous effort with no division of the task. This task centred perspective may be appropriate when the focus is on task completion but other issues become important if the emphasis is on learning outcomes, or on the process of collaboration itself.

Whether collaboration is the means to the end of learning some domain material or whether collaboration is in some sense the end itself. In the former case, learning efficiency may be most important but in the latter case, efficiency is not the key issue — indeed, when learning to collaborate is the primary goal it may well be that inefficient domain learning is inevitable.

One view of collaboration is that part of the process of task performance requires that there is some agreement about the nature of the task — and this can be quite complex. For example, the view of Roschelle and Teasley that collaboration is "a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem" (Roschelle and Teasley, 1995). For Roschelle and Teasley, part of task performance involves synchronous effort with no division of that particular part of the task. This approach may serve to define when collaboration is taking place and when it is not — but it provides little help in determining, for example, the quality of the collaboration or how learners become more proficient at collaboration.

In contrast, here, learners are assumed to have some task to perform together, and that some way is required of estimating the quality of any perceived collaboration. The emphasis is therefore on the collaborative process as opposed to the collaborative state. The main issue that is examined is learning to collaborate so domain-based learning outcomes are of less interest.

The key notion is that good collaboration is characterised by an even distribution of dialogue roles. The justification is based on a model of distributed cognition in which cognitive processes communicate with each other (Minsky, 1987). The further assumption is that learning to collaborate is best for the group of learners when each learner exercises all the dialogue roles available, and all dialogue roles are exercised equally.

Defining dialogue features as dialogue roles, and analysing dialogue for these roles gives a broader framework in which to view collaboration. This is useful since Webb and Cohen (and others) come up with different necessary conditions for good collaboration (Webb, 1983; Cohen, 1994). For example, while Cohen suggests that the task may be responsible for determining what roles are possible in a situation, she also argues that roles (and dialogue features) are crucial in themselves.

Policies for Collaborative Activity

The aim is to examine which policy is likely to be "best" in terms of some measure. To achieve this, a number of likely candidates are described. The main hypothesis investigated is that "Normal" collaboration, unrestricted by any enforced distribution of dialogue roles[1], except that certain social norms operate, does not yield the optimum results for learning to collaborate (according to the measure we adopt below).

The basic policy that is closest to perhaps the most familiar form of collaboration is termed *Free*. For this

[1] i.e. linguistic functions of utterances used in collaboration

"normal" case, agents are permitted to drop roles whenever they wish. The primary restraint on their role usage is that two agents are not permitted to have the same role at the same time — roughly, normal social behaviour.

Other forms include *Multi*, *Swap* and *Polite*. Viewed as constraints on behaviour, some of these forms can be combined to yield, for example, *MultiSwap* and *MultiPolite*. The *Multi* policy allows agents to use the same role at the same time (to do so, the number of roles had to be slightly simplified). The resulting conversations will be "strange" in that the dialogue will seem to lose its coherence to some degree as questions are asked together, and agents may "reply" with other questions (Clarissa agents have a very good memory which is a key feature of this collaborative situation).

For the *Swap* policy, the opening of a new dialogue game is an indication of a new episode, and at this point, agents drop the roles they are playing (it does not follow that they would choose different ones for the next episode). This approach is derived from an observation by Soller who found that in the cases where people do choose different roles at the beginning of new episodes, collaborative activity is more beneficial (according to her measures) (Soller, 1997). The *Polite* policy arranges for agents to 'drop' their roles at the *end* of episodes. Rather than swapping roles at the beginning of a new episode, the participant who has lead an episode stands back for the next. In other words, if you have been taking the lead for a while, stop, and let somebody else take the floor.

MultiSwap, in an educational context, is equivalent to a situation in which collaborators can ask all the questions they have, and say everything they can about a problem. They pay careful attention to everything said by their partners. Questions, comments and suggestions must be noted down. An alternative approach is *MultiPolite* collaboration which differs from *MultiSwap* in the same way that *Polite* and *Swap* differ.

Clarissa's Architecture

The architecture is divided into two main subsystems: the cognitive and dialogue systems. The cognitive system is broken into separate units (cognitive processes), which then communicate with each other.

This architecture is rooted in notions of distributed cognition. Hutchins draws a parallel between the division of labour and the division of cognitive processes (Hutchins, 1995). Collaborative activity may not be obviously divided at the domain level, but individuals may each be exercising different cognitive processes. Clark and Chalmers advocate that cognition is not bounded by the skin, but encompasses a wider context (Clark and Chalmers, 1998). They develop active externalism, noting that internal cognitive processes are coupled in a two-way interaction with external entities. Bearing in mind that Vygotsky talks in terms of dialogue behaving as stimuli in the process(es) of solving a task, Clark and Chalmers go on to discuss the function of language as the means of coupling cognitive processes to the external environment. They conclude that the ability to communicate between internal and external processes, extended cognition "...is no simple add-on extra, but a core cognitive process". Here, this is taken as meaning that the mechanism by which processes are coupled (and communicated) internally are very similar to those which allow external coupling.

All communication that might normally be regarded as occurring within the cognitive system are made using the dialogue system (see below). The decision about which utterances are 'hearable', and which passed directly between cognitive processes within the agent are made within the dialogue system. Other than with respect to communication, individual cognitive processes may be implemented in any way. Clarissa itself uses a very simple forward chaining production rule system.

The dialogue system uses a 'dialogue game' mechanism to maintain the focus[2], and form a prediction about what might be said next. A dialogue game is defined to be a state machine which represents the entirety of possible dialogue utterances and the order in which they can occur. Parallel dialogue threads are used to let agents keep a number of topics active at the same time.

The dialogue role mechanism is used to control the communication between cognitive processes. A dialogue role can be seen as defining a (complex) zone within a dialogue game, a 'sub dialogue game'. Roles that one agent is playing cannot be used by another. Roles are swapped frequently, and the effect of this restriction is to model the way people normally respond to each other, one asking a question, the other replying and so on. To decide what is said next, both the role that individuals are playing, and the dialogue game is examined. Clarissa allows for a variety of dynamic mechanisms for distributing roles throughout the period of the ongoing dialogue.

The dialogue system must allow messages to be passed between the cognitive processes found in the cognitive system. To achieve this communication, a new abstraction is introduced, a *dialogue goal*. A dialogue goal is an expression of the communicative act (Maybury, 1993) that the cognitive system wishes to take place. The dialogue system can then choose how best to achieve this goal. The cognitive process initiates a dialogue goal whenever it wishes to pass a message to another process. These are 'thrown over the wall' to the dialogue system which requires that the goals the cognitive system generates are understandable by the dialogue system so there will be some dependency between these two systems. Goals so delivered, and acted upon by the dialogue system may eventually be completed. The desired result is that the relevant messages are delivered back to the cognitive system. To do so requires that there is some mechanism in the dialogue system for passing information back to the cognitive system.

Any number of Clarissa agents can collaborate, and they do so within a specific computational context. The task context used for this paper requires that the simulated novice physics students have to draw a diagram to

[2] Related to McCoy's approach (McCoy and Cheng, 1991)

represent how energy flows through a n electrical circuit with a charged battery connected to a bulb (with two wires), and makes a bulb shine. Space precludes a full discussion of the architectural issues — see (Burton, 1998; Burton et al., 2000) for more details.

Experimenting with Clarissa

The Clarissa system was set up to run two Clarissa agents collaborating on the standard test problem for a hundred runs for each of a number of policies (Clarissa agents act in a distributed network). The results were interpreted in terms of the distribution of roles throughout the task performance. This is in line with the underlying claim that exercising the range of available roles is advantageous, and that role usage is balanced between the participants.

Role usage is approximated by the number of utterances which are made by participants while an individual is 'playing' that role. Different policies are compared so that a provisional determination can be made about the ways in which collaboration can be organised to yield good collaboration. To analyse the various different collaborative environments samples of 100 pairs were used. Table 1 presents the important statistical measures.

There are two key ideas in the (non-standard) statistical analysis. Firstly, there is an analysis of the degree to which agents are split by the collaborative environment within a pair, so that one performs differently from the other. Secondly, it is also possible to discover something about how they perform differently by looking at the correlation between the agents. It might be hoped that agents in a pair perform equally. The above statistic (the significant split) will tell us whether it is likely that the differences between the participants in a pair can be accounted for simply by the variation in the population in general. Ideally, agents in a pair that uses a given role a lot shares this usage between the agents. In other words, if one agent uses a role a lot, the other should do likewise. For the purposes of this study, it is desirable to have a positive correlation between the two agents for all of the roles investigated. Negative correlations are likely to result in significantly split role usage. A negative correlation implies that the collaborative environment, of itself, is inducing the agents to divide the roles unevenly.

Any role usage which seems to be negatively correlated between two agents is of interest. Positive correlations are to be encouraged, as they imply that the agents are splitting the role between themselves sensibly. These statistics are used to determine how many of the roles are being played equally by participants in a pair, and how many are unevenly split, for a given collaborative environment. Additionally, the number of times agents receive interesting information from their partner is used[3]. In summary, the metric for good collaboration, according to the criteria that roles should be evenly distributed is that: Neither roles, nor the number of times agents receive interesting information from their partner, should be significantly split, or negatively correlated.

In Table 1 the results are interpreted in the case of the *Free* policy, 5 out of the 7 roles, and the number of "interesting" moves received by agents (6 out of 8 in other words) are significantly negatively correlated. this statistic indicates the degree to which the two agents in the pair share out the usage of a role.

In this case, as one agent used a role more, the other uses it correspondingly less, (and similarly the number of interesting events recoded by each agent follows a similar pattern). Correspondingly, all those roles which are significantly split, are also significantly negatively correlated (again, at the 1% level). Here the statistic is measuring the propensity of an agent to use a role more while their partner uses it less. Thus in the *Free* case, 6 out of 8 roles are negatively correlated, indicating that one agent tends to use the role more while the other uses it less.

The effects of this collaborative environment can be characterised in terms of the measure defined above as having 6 out of 8 roles significantly split and negatively correlated and a mean number of interesting events of 24.9 which is significantly split between agents in a pair, with the mean difference being 12.8.

The result of this experiment suggest that if all other factors are equal, students that collaborate together, with no control on their dialogue, are likely to benefit unequally from the experience. The first student who takes control of the conversation will remain in control, while the other participants adopt a more subservient role. This is undesirable from the perspective adopted here, but possibly common in small group work.

Examining the *Swap* policy, the reason role for the *Swap* policy is still significantly split to roughly the same degree as for the normal environment, but overall the picture is very different (Table 2). Now only 2 out of the 8 roles (and events) are significantly split while 5 are negatively correlated. The statistics suggest that it is no longer possible to be certain (at the 1% level) whether there is a significant split between agents in the pair. But in this case the mean of the difference between the agents is still relatively high (11.9) and fairly similar to the previous environment (12.8). The difference here is in the spread of the original population.

The correlation coefficient tells a similar story. It is worth noting at this point that in the case of the "interesting" events, 8 outliers have been removed from this data sample (by the procedure described above). If this had not been done, the resulting correlation would be about 0.614, significantly positive. However, in the overwhelming majority (92%) this would be misleading, as in their case the correlation is calculated as -0.462, significantly negatively correlated. The interest here is in the majority, so it is important to remove the outliers. The figure of -0.462 is still significant and leads us to believe that this sort of distribution would not have happened by chance, but it is clearly less so than the -0.99 seen in the previous example.

In common with the "interesting" case above, a number of the roles are split, but not significantly so (at the

[3] In this work, "interesting" information is information that leads to an agent's knowledge base changing.

Table 1: Results for the *Free* policy (**Bold** entries statistically significant)

	Interesting	check	reason	generate	response	interface	argue	question
Correlation	**-0.972**	-0.189	**-0.99**	**-0.99**	**-0.765**	**-0.985**	**-0.974**	-0.11
Split	**11.10**	1.41	**19.75**	**19.69**	**3.12**	**16.13**	**12.39**	1.28

Table 2: Results for *Swap* environment (**Bold** entries statistically significant)

	Interesting	argue	check	generate	interface	question	reason	response
Correlation	**-0.462**	**-0.56**	-0.19	**-0.58**	**-0.281**	-0.0105	**-0.594**	0.186
Split	2.22	2.39	1.35	**2.67**	1.68	1.13	**2.68**	0.96

1% level). They are on the borderline, and are rejected by our relatively stringent test. We expect that if roles are negatively correlated, they are also likely to be significantly split. In this case, while the correlation statistic yields a value which is significant at the 1% level, a number of the "split" statistics are not significant at this level. The implication is that the correlation between the two statistics is not as linear as might have been suspected. The fact that roles are significantly negatively correlated, but not significantly split means that both statistics need to be considered. The interpretation of such a scenario is that in this environment, the inclination of participants is that if one starts to use the role more, the other will use it less. However, this happens to a lesser degree than the overall distribution of the way in which participants use the roles anyway and (at this given level of significance) they do not divide the way in which they use the roles.

In conclusion, the *Swap* environment does indeed produce somewhat better collaboration according to our measure. This finding indicates that our first hypothesis is correct. Simple, unconstrained collaboration is not necessarily the most effective. But the *Swap* constraints still produce an environment in which a number of roles are split between participants, and many which are negatively correlated. In short, *Swap* is better, but not best! The mean of the number of interesting events found by the agents has also risen, and, as seen above, the degree to which the level of interest is split between partners in a group has fallen. However, the rise is less than our chosen statistical degree of significance, so the idea that this has happened by chance is not rejected.

Multi, the next policy to be examined, is based on allowing participants to put forward several ideas in parallel, not necessarily commenting on any of them. In practice, each participant may ask several questions, and their partners may not necessarily answer them directly (and likewise for other dialogue utterances).

It is important to remember at this stage that the Clarissa agents have infallible memories. If this form of collaboration were to be attempted between human participants, not only would the participants need to be trained in the dialogue structure (or forced to use it by an interface), but they would also require some assistance in terms of remembering what had been said.

In this case, neither the interesting events, nor the reason role, are significantly split (Table 3). However the interesting events show a strong negative correlation. Again, the pairwise difference being not significant in this case implies that while there is a tendency for one participant to receive more interesting events from their partner than vice versa, the degree to which this happens is less than the distribution of how interesting events are seen by participants anyway.

The only other role to be negatively correlated is also significantly split. That is the "interface" role. Effectively, this is the role used by participants wishing to use the mouse. In many respects we would guess that this would be the most difficult role to make sure was evenly divided between the participants.

The complete metric for the *Multi* policy is measured across 6 roles (including the "interesting" events). In this case, "check" has been amalgamated with "question" and "argue" with "response". These simplifications were made to keep the complexity of the dialogue game definition within a reasonable level (since all roles must be replicated for *Multi* and all replicated roles must be related to all others). The results are 1 out of 6 have significantly split role usage, and 2 have negative correlation. The mean of the interesting events is 26.6. Again, this mean is not significantly different from either of the means of the other two samples.

The collaborative environments which *Multi* represents seem to have allowed agents to distribute the roles more evenly, but at the cost of relying on agents having a perfect memory for what has been said by others, (so that they may return to these previous statements). The degree to which roles are more evenly distributed in this case could was not predicted. It seems that allowing agents to play the same role together allows them to avoid situations in which one agent simply takes control. This is a useful finding.

The model seems to be implying that it is not necessarily the case that the individuals should be encouraged to practise different collaborative skills than their partners (at any one time), but that it is both acceptable and possibly beneficial at least to allow them to practise sim-

Table 3: Results for the *Multi* policy (**Bold** entries statistically significant)

	Interesting	generate	interface	question	reason	response
Correlation	**-0.649**	-0.8	**-0.759**	0.121	0.332	0.043
Split	2.59	0.39	**3.36**	1.07	0.80	1.10

ilar skills concurrently. A rise was noticed in the interest level shown in *Swap* collaboration above, which was not significant at the 1% level. The next step is to examine a policy which exploits this rise, in conjunction with the better role usage seen for the *Multi* policy.

Both the *Multi* and the *Swap* constraints can be combined, since they are addressing different aspects of the environment. The former is about the roles which can be used together at any one time, the latter is about the points at which agents are expected to swap roles. It is not so clear that by combining them the results are positive. Not only are the results positive, taking the best from both individual schemes, but that they are better than one would have imagined (Table 4). Neither of the two data sets show any degree of split or negative correlation, and this is common across all 6 data sets, including the "interface" role which in the "multi" environment was still split. In the case of the "reason" role there is a degree of positive correlation (which is significant at the 1% level, as it is greater than 0.254). This is highly desirable as it implies that one participants uses the role more, the other will follow suit.

Furthermore, the level of interest is higher than seen in the "multi" environment. This must be discounted as the difference is very small and statistically could easily have happened by chance. Indeed the probability of the rise in the degree of interest from the *Free* case (24.9) and this case (29.0) happening by chance is still about 27%, a very long way from the 1% cut off level.

Nonetheless, this mode of collaboration is clearly much better according to our metric, especially in so far as it seems to encourage the even distribution of all dialogue roles. This implies that all the participants in the collaboration will have an opportunity to practise and observe all of the dialogue roles that are available, and by doing so, if these roles have been chosen to reflect underlying cognitive processes which are pedagogically interesting, the participants should have the opportunity to practise, observe and improve their ability to execute those underlying processes.

It has been our goal to find such a cognitive environment. Having done so it is necessary to reflect on how this environment relates to a realistic situation with human collaborators. It was previously noted that the human participants would need some additional assistance to allow them to keep track of what everybody had said. (This is common in most meetings, and often some such device, like paper and pencil, is encouraged in educational collaboration). They also need to be encouraged to speak their mind, ask all their questions, say everything they can about a problem, with, perhaps little regard for their partners. In another respect, they must of course take careful note of everything their partners say, such that they can respond at some point. This satisfies the "multi" constraints. Finally the *Swap* constraints imply that participants should keep asking questions, if that is what they start doing; keep making statements about what is known about the problem; keep proposing deductions, suggesting things to do next. Participants should continue concentrating on one "role" until the beginning of a new "dialogue episode". The beginning of a new dialogue game has been used as such a marker. The key is to guarantee that these events happen frequently enough that participants do not get frustrated and bored by being captured in one role for too long a time.

Of course it is necessary to re-emphasise that this is the finding of a simulator which is built in a cognitively plausible manner, but has necessarily simplified many issues. The results of simulated model can do no more than suggest ways in which others may take this research further, having decided upon their pedagogic requirements, and the dialogue roles that they hope will fulfill these.

An alternative approach to the *Swap* form of role swapping, as has been mentioned previously, is the *Polite* mode. Here, rather than everybody swapping roles at the beginning of dialogue episodes, participants who finish dialogue episodes agree to drop their roles, allowing somebody else to take the floor. This may, in some situations, be an easier form of dialogue to explain and encourage in participants. Simply, having led the discussion about a specific topic, a participant should allow their partner to lead the next. This behaviour can be combined with the *Multi* behaviour previously examined (*Polite* is very similar to *Swap* so it is not presented here).

The findings, shown in Table 5, for this mode of collaboration are similar to those of *MultiSwap* collaboration looked at above. None of the roles are significantly split or negatively correlated. The mean number of interesting utterances received by one partner from the other in this case has risen to 33.9 (s=8.41). The weighted standard deviation of this population and the population of *Free* collaborative partners (s=1.63) is 8.57 (2 d.p.). The difference in the mean is 9, which is just 1.05 SD. This is still not significant at the 1% level, indeed it would be expected in 29.38% of normally distributed cases which had similar means. Little can be said about the increase in the level of interest seen, except to note its increase, and to note that it is slightly above that for *MultiSwap*.

There is then very little to choose between the *MultiSwap* environment and the *MultiPolite* environment. Both yield good collaboration as measured by our metric. The decision about which to use may rest on the ease

Table 4: Results for *MultiSwap* environment (**Bold** entries statistically significant)

	Interesting	generate	interface	question	reason	response
Correlation	-0.238	0.874	-0.223	0.0662	0.632	0.0787
Split	1.53	0.30	1.45	1.00	0.47	1.11

Table 5: Results for *MultiPolite* environment (**Bold** entries statistically significant)

	Interesting	generate	interface	question	reason	response
Correlation	0.194	0.691	0.0567	0.128	0.666	0.0684
Split	0.93	0.45	1.10	1.01	0.50	0.98

with which the various different behaviours demanded by the collaborative environment can be taught to the participants or engineered into the environment in which they are collaborating.

Discussion

Clarissa agents simulate a form of collaboration. They offer a test-bench for investigating collaborative activity. Various different policies for collaboration have been, and others could be, examined with Clarissa. Different opinions about what makes for good collaboration can be used to investigate a variety of different collaborative situations. The notion that good collaboration is characterised by an even distribution of roles has been adopted and some hypotheses about collaboration have been evaluated given the starting point that even role distribution is important. The following results have been found:

- The best collaborative environment found involves participants having the ability to communicate in a non-normal way, and having some form of aide-memoire which would allow this abnormal communication to be effective. This finding was not predicted.
- Socially "normal" role usage is not conducive to educationally beneficial collaboration. One agent takes control of the conversation, and the social norms are such that the agent remains in control. This problem seems to be addressed by the multi-role distributions. Allowing participants to adopt the same roles at the same time is not socially normal, but seems to have a very significant effect on the quality of collaboration.

In the best form of collaboration that has been found, the environment involved participants swapping roles frequently. They can be instructed to do so either when they finish, or start, a new dialogue episode. The simplest instruction may be "if you have been dominating the discussion, when you come to the end of a topic, allow somebody else to take the lead". Clarissa has given interesting results that can be taken further.

Acknowledgments

Mark Burton was supported in this work through an EPSRC PhD Studentship.

References

Burton, M. (1998). *Computer Modelling of Dialogue Roles in Collaborative Learning Activities*. PhD thesis, Computer Based Learning Unit, The University of Leeds.

Burton, M., Brna, P. and Pilkington, R. (2000). Clarissa: A laboratory for the modelling of collaboration. *International Journal of Artificial Intelligence in Education*, 11(2):79–105.

Clark, A. and Chalmers, D. (1998). The extended mind. *Analysis*, 58(1).

Cohen, E.G. (1994). Restructuring the Classroom: Conditions for Productive Small Groups. *Review of Educational Research*, 64(1):1–35.

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press, Cambridge, MA.

Maybury, M. (1993). Planning multimedia explanations using communicative acts. In Maybury, M., editor, *European Conference on Hypertext' 94*. MIT Press, Cambridge, MA.

McCoy, K. and Cheng, J. (1991). Focus of attention: Constraining what can be said next. In *Natural Language Generation in Artificial Intelligence*, chapter 4, pages 103–124.

Minsky, M. (1987). *The Society of Mind*. MIT Press, Cambridge, MA.

Roschelle, J. and Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In O'Malley, C., editor, *Computer Supported Collaborative Learning*, pages 69–97. Springer-Verlag, Heidelberg.

Soller, A. (1997). Personal Communication.

Webb, N.M. (1983). Predicting Learning from Student Interaction: Defining the Interaction Variables. *Educational Psychologist*, 18(1):33–41.