Learning to solve syllogisms by watching others' learning*

Padraic Monaghan and Keith Stenning

October 2, 1998

Introduction

This experiment arose from the confluence of two interests, one in "vicarious" learning (learning by observing others learning) (see e.g. McKendree, Stenning, Mayes, Lee, & Cox, 1998), and the other in the learning of elementary logic (see e.g. Stenning, Cox, & Oberlander, 1995; Stenning & Yule, 1997; Monaghan & Stenning, 1998). The experiment involves teaching students a method of solving syllogisms using Euler's Circles. The teaching session includes "vicarious" learning through video clips of students solving syllogisms with this method.

The study addresses several questions:

- how do students respond to graphical methods of problem solving?
- what are the individual differences in students' abilities to acquire and utilise a graphical method for problem solving?
- does "vicarious learning" aid acquisition and use of the graphical method?
- what are the effects of watching a fellow student struggling to use the method and a student using the method with ease?

The motivation for using Euler's Circles came from several directions. Syllogisms are complex reasoning problems that evince marked differences in solution strategy and the ease with which students answer them. Therefore, there is a great deal of variation in the way the problems are approached and responded to. Syllogisms can be solved using a graphical strategy – that of Euler's Circles – and, again, students indicate variations in the ease with which they acquire and utilise this strategy (Monaghan & Stenning, 1998). Due to its being a graphical strategy, it is a demonstrable method. Students use external representations of the graphics in order to solve syllogisms, therefore seeing the method in operation as performed by another student is clear.

Method

Subjects

The subjects were 32 undergraduate students from the University of Edinburgh. 30 of these were undertaking the experiment as part of a course in human communication, 2 students participated in order to fulfill criteria for a linguistics course.

*This research was supported by the ESRC Cognitive Engineering Programme, the EPSRC Multimedia and Networking Applications Programme, and by ESRC research studentship No. R00429634206.

Materials

The teaching was performed in one hour sessions, and involved a pre-learning and a post-learning test. The students were asked to watch a video which gave an introduction to syllogisms[1]. They were then required to answer 2 practice syllogisms, and were given an opportunity to ask questions. Following this they were given ten syllogisms to answer in 10 minutes and asked to pace themselves so that they attempted most questions. There were 8 syllogisms with valid conclusions and 2 with no valid conclusion. The syllogisms with conclusions were selected such that they increased in difficulty, according to the findings of Johnson-Laird (1983), and Ford (1995).

For the second part of the experiment, all students were instructed by video in the Euler's Circle method for solving syllogisms. The method was summarised on the front page of the booklet containing the syllogisms in order to reduce memory load. The summary of the method read as follows:

- translate the premisses into diagrams;
- superimpose the B-circles, and maximise the number of regions;
- keep a cross if its area has not been divided up;
- make your conclusion (if there is one).

Once the method had been described, two examples of the strategy used to solve syllogisms were presented to the students. Students were then given the chance to ask the experimenter questions.

At this point, 3 groups were separated. The control group (n=10) went immediately on to attempt to answer the syllogisms. The second group (n=11) (the "struggler" group) saw video clips of a student struggling to answer the two syllogisms that had been used as examples of the method for all groups. The third group (n=11) (the "sailer" group) saw video clips of a student answering with ease the two syllogisms that had been used as examples previously. Transcriptions of the video clips are given in Appendix 2. Following the video clips, the students were asked to attempt the 40 syllogisms in the booklet. There were 26 syllogisms with valid conclusions (i.e., all the syllogisms with valid conclusions bar the one used in the instruction materials) and 14 syllogisms with no valid conclusion. There are 21 different forms of integrated diagrams that can be constructed using the Euler's Circle method, of these 8 render valid conclusions. The syllogisms were ordered such that each form of integrated diagram with a valid conclusion occurred at least once early in the set of problems.

The students were informed that there were probably too many problems for the time they had but that they should try and work at the same pace as they did in the first part of the experiment. They had 20 minutes to attempt the questions.

Results and Discussion

For each group, the number correct on the initial 10 syllogisms was scored ("pretest"). For the post-intervention test, the number correct on the first 10 ("post10"), the first 16 ("post16"), and all 40 syllogisms were recorded. The first 16 was chosen as this was the maximum number of syllogisms that all students attempted. The table indicates the mean scores for these measures:

[1] a transcript of the video materials is found in Appendix 1.

group	pretest	post10	post16	postall	n
all	3.75	4.59	8.28	11.41	32
control	3.50	4.90	8.70	16.30	10
struggler	4.55	4.64	8.27	9.64	11
sailer	3.18	4.27	7.91	8.73	11

Scores were analysed using one-way ANOVAs with group as the factor, and score as the dependent variable. The only significant finding was in the postall score: $F(2,29) = 3.72$, $p<0.05$. the Scheffe post-hoc test indicated that the significant difference is between the control and sailer groups. Hence, the control group scored higher than the group that watched the "sailer" student. This was an effect on the speed that the students worked at, the number correct being a proportion of the number attempted: a one-way anova with number of problems attempted as the dependent variable and group as factor gave $F(2,29) = 21.02$, $p<0.0001$, with the Scheffe test indicating that significant differences are between the control group and both the "vicarious" groups.

This illustrates that watching other student's learn has a large effect on pacing. Those students that did not see a student solving the syllogisms worked the fastest. Those students that watched a video of a student solving syllogisms worked at a pace more in keeping with the experimental instructions, and more akin to the pace of the students they watched solving the problems (though not quite so slow).

Another question addressed in this study is whether certain groups indicate more improvement than others. In other words, what is the relationship between the teaching materials and scores on the post test as compared to the pretest?

Scores on post10, post16, and postall all correlated significantly with scores on the pretest ($r = 0.70$, $r = 0.70$, $r = 0.57$, respectively, all $p<0.001$). However, pretest score did not correlate with number of syllogisms attempted: $r = -0.06$, $p>0.7$. Correct attempts before being given a strategy related to the success of the strategy being acquired, but does not seem to relate to the speed of its use.

Do lower-scores and higher-scores on the pretest indicate different responses to the different teaching materials? In order to assess this question, the groups were split along the median of scores on the pretest. The following table indicates scores for the groups split along these lines:

group	pretest	post10	post16	postall	n
all	3.75	4.59	8.28	11.41	32
control low	2.00	4.00	7.40	10.60	5
control high	5.00	5.80	10.00	22.00	5
struggler low	2.50	3.00	5.25	5.50	4
struggler high	5.71	5.57	10.00	12.00	7
sailer low	2.14	3.43	6.57	7.29	7
sailer high	5.00	5.75	10.25	11.25	4

None of the differences between low-scoring groups were significant, so students who struggled were not assisted by any of the teaching materials in particular. Equally, none of the differences between high-scoring groups were found to be significant, though the high-scorers who watched the student struggling scored slightly less on the post10 questions than they did on the pretest. This reduction may well be a ceiling effect, however, as the high-scorers in the "struggler" group scored the highest average on the pretest, and scored comparatively with the high-scorers in the other groups on the post-test.

Results summary

These initial analyses indicate that the general effect on learning of seeing other students learn was solely to do with the pace at which students performed their task. No significant differences in ability to use the Euler's Circle strategy were observed after taking into consideration the effect of pace of response.

One avenue of investigation remains: do students learn different things from the different materials? We have seen no quantitative differences in scores, but are there qualitative differences in the type of errors made, or the type of questions that one group get right? What are the individual differences in response by the student to the teaching materials? The next section considers these issues.

Individual Differences

In order to raise hypotheses about how students may learn differently it is important to consider in detail the teaching materials. All students have the same instructions, but some students see extra examples of learning. What instances of learning may arise from these video materials? It is necessary to know what errors are made by the student and what promptings are made by the instructor in order to assess the implications of these materials.

Struggler

- syllogism *all architects are botanists, all botanists are chemists*
  - the student does not draw premiss diagrams separately initially.
  - the student has difficulty with the cross from the second premiss, tries to place it in the A B C region, in the not-A B C region, and on the line of the B circle.
  - the student is instructed to redraw the second premiss.
  - the student is instructed that the B circle is split by the a circle, he does not perceive that the A circle divides up the area.
- syllogism *all architects are botanists, some botanists are chemists*
  - the student is instructed to put the B circles on top of each other.
  - the student is instructed to maximise the number of crossings.
  - the diagram drawn by the student is slightly confusing, creating two areas that represent the same groups: not-A B not-C.
  - the student is asked about the cross from the second premiss. the student asks whether it disappears because it crosses over the A circle.
  - the student is asked to indicate the region some Bs are Cs in the diagram.
  - the student is told that this area (some Bs are Cs) is split.
  - student makes wrong conclusion "some chemists are architects", and asked why. replies that it's because the C and the A circle overlap. the student is then instructed that there is no information that this region contains anything.
  - the student has further difficulties making a conclusion: says "not all ..." before correctly answering "no conclusion".

Sailer

- syllogism *all architects are botanists, all botanists are chemists*

    - the student delays slightly in making a conclusion - aware that the cross from the second premiss requires thought.

- syllogism *all architects are botanists, some botanists are chemists*
    - the student draws the correct diagram, and talks through the removal of the crosses.
    - the student has a slight problem with making a conclusion: "I seem to have got rid of all my crosses, so I don't know." The student is instructed that if there are no crosses left then that means "no conclusion."

The students whose proofs were watched in the two video conditions indicated difficulties at different stages in the procedure. In order to assess the effects (either positive or detrimental) consequential upon these video materials it is necessary to distinguish different types of error in the written performance of part 2 of the experiment. Four types of error were distinguished:

- errors of translating into the premiss diagrams (tran in);
- errors of arranging the circles into an integrated diagram (diagram);
- errors of placing crosses into the integrated diagram (crosses); and
- errors of translating from the integrated diagram to give a conclusion in sentential form (tran out).

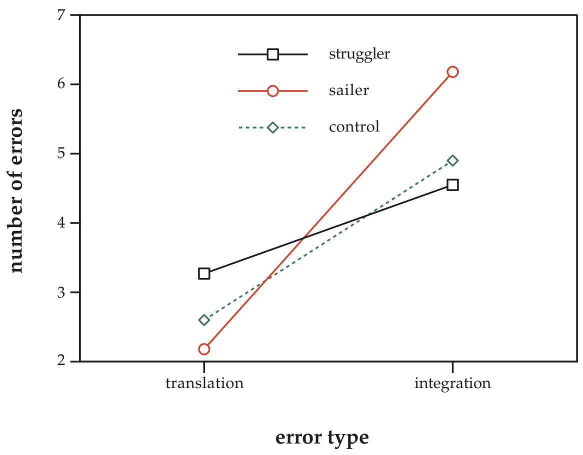
Only errors that led to the student producing the wrong answer were counted, so if a student gets the wrong integrated diagram but produces the correct sentential conclusion this does not count as an integration error. The earliest error in the procedure only was recorded, so if a student gets both the arrangement of the circles *and* the position of the crosses in the integrated diagram wrong, for example, only the arrangement of the circles counts as an error. The first 16 syllogisms of part 2 were analysed.

We may anticipate that the students that watched the struggler may have more assistance in constructing integrated diagrams and placing crosses in that several errorful possibilities were gone through in the course of the video - the sailer's performance provided no cues for "what not to do". In the group that watched the struggler, more errors may be expected in translating into the diagrams, as the method was not clearly indicated by the struggler - he produced an integrated diagram before drawing the premisses separately, but this stage in the strategy was performed systematically by the sailer. Hence, watching the student that struggled may have a detrimental effect on translating from the premisses and a positive effect on producing the integrated diagram. Watching the student that had little difficulty ought to be beneficial to translating from the premisses, but have little effect on producing the integrated diagram.

The following table shows the mean number of errors of each type according to group (tran is the sum of tran in and tran out errors, int is the sum of diagram and crosses errors).

group	tran in	diagram	crosses	tran out	tran	int	proportion translation	n
all	1.00	3.47	1.75	1.69	2.69	5.22	34%	32
control	1.30	3.70	1.20	1.30	2.60	4.90	35%	10
struggler	1.45	2.64	1.91	1.82	3.27	4.55	42%	11
sailer	0.27	4.09	2.09	1.91	2.18	6.18	26%	11

The table and graph suggest that students who watched the struggler make fewer errors in creating integrated diagrams. Those that watched the sailer made fewer errors on translating. This offers (non-significant) support for our hypotheses regarding what a student learns from the



video material. A result of note is that students in the sailer condition on average made more errors in constructing the integrated diagram than did the control group. Hence, the video may have been detrimental in this respect, though the difference is non-significant.

Though the table and graph suggest that subjects learned different skills from the video, a two-way repeated measures anova with error type as the within-subject factor and training condition as the between-subjects factor revealed only a significant effect of type of error ($F(20,1) = 9.21$ $p<0.01$). There was no significant effect of training condition, and no significant interaction of error type and training condition ($F(20,1) = 2.46$, $p = 0.132$). The only difference that approached significance was between the two condition groups for tran in errors: $t(20,1) = 1.74$, $p<0.1$.

When question type (with/without valid conclusion) is distinguished, then there is a significant interaction of error type and training condition for questions with no valid conclusion ($F(20,1) = 6.26$, $p<0.05$). This interaction is in the same direction as the above graph. This is perhaps due to the lengthy learning process that the struggling student undergoes on the video for the syllogism with no valid conclusion. The students watching the struggler learn more about using the Euler's Circles method for the no valid conclusion cases.

The possibility remains that the effects of vicarious learning could occur at a much more local level than in the general application of a strategy. Do the different groups respond differently to particular questions? The following table indicates the number correct and the error type for each of the first 16 syllogisms from part 2[2].

[2]The syllogism notation uses small letters for quantifiers ("a" = all; "s" = some; "n" = no) large letters for properties (A, B, or C) and "-" for not. Where the columns do not add up there were students who did not provide a conclusion for a particular question.

syllogism	correct			error type					
	control	struggler	sailer	control		struggler		sailer	
				tran	int	tran	int	tran	int
nBAaBC	0	1	0	5	5	5	3	3	7
sBAaBC	6	7	3	2	1	1	3	4	3
sBAsCB	7	9	8	1	2	0	2	0	3
aBAaBC	4	6	2	0	5	1	4	8	1
sABnCB	2	1	2	3	5	4	6	2	6
sB-AaBC	3	2	3	4	3	6	3	5	3
aBAaCB	**8**	**9**	**10**	**2**	**0**	**3**	**0**	**1**	**0**
aABsC-B	5	4	5	3	2	3	4	0	6
nABsB-C	5	5	4	0	5	3	3	0	7
aABnCB	9	8	9	1	0	2	1	2	0
aABaCB	7	6	6	0	3	1	4	0	5
sBAnBC	3	2	1	1	6	4	5	3	6
aBAsCB	7	5	8	1	2	1	5	0	3
aABsCB	**7**	**10**	**7**	**0**	**3**	**0**	**1**	**0**	**4**
nBAaCB	9	8	10	1	0	2	1	1	0
sA-BsB-C	5	10	8	0	5	0	1	0	3

Certain syllogisms render only one sort of error for particular groups, but how can they be explained, and are these results significant? There is an isomorphism between the Euler's Circle diagram for the syllogism aBAaCB (this is highlighted in the table above) and the training syllogism aABaBC. However, most students in all three groups solved this successfully, and any errors were due to translating difficulties. There is also an isomorphism between the Euler's Circle diagram for the syllogism aABsCB (also highlighted above) and the training syllogism aABsBC. The students who watched the struggler on video were most successful as a group in solving this syllogism. They learned the solution for this type of syllogism well, even though the application of the strategy to other problems was not generally better. Errors on this syllogism were all errors in the construction of the integrated diagram, the video material of the struggler presumably assisted in the construction of this sort of diagram.

The nature of these differences in performance on particular questions remains a topic of further research.

Conclusions

The main findings of this study are that vicarious materials seem to have an influence on pacing during task performance – students that watch another student solving problems pace themselves according to the speed with which the problems are solved in the viewed material. Also, it seems that different materials corroborate different aspects of the method. Watching a student who has difficulty and dwells on creating the final integrated diagram in the method seems to lead to better performance on creating the integrated diagram, particularly on questions of a similar type to the problem viewed. Watching a student who systematically and clearly translates the premisses in and out of graphical form seems to benefit students in the translation parts of their application of the strategy.

So, different materials teach different aspects of a strategy. However, there were no aspects of performance where students who watched other students learning did significantly better or worse than the control group. Why would this be? There are several possibilities as to why the extra material had no appreciable effect. It may be that watching other students learning really does not benefit the student to acquire and apply the strategy in this case. Other possibilities are that the delay between the initial instructions and the chance to apply the strategy was critical for the groups that watched the extra video material. This may have over-ruled any beneficial effect of the vicarious material. One student from the struggler group commented that all the errors that the videoed student made got in the way of his understanding. Perhaps the extra material was too short, or too limited in the problems it contained. One student from the sailer group commented that it would have been more useful if the video had contained some "harder problems". Another possibility is that students who watch other students learning may be negatively influenced by the errors the student makes, or by the individual style of that student. Perhaps the sort of errors viewed are critical, and some may be beneficial, whereas other's may be detrimental.

Though largely failing to reach significance, the results of this study suggest that subtle differences in the video material that students view lead to variations in performance on different stages of a taught strategy. These variations, properly controlled, may be used to aid use of certain aspects of a complex strategy, and used to strengthen and compensate for the weak areas in a student's abilities.

References

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition* 54, 1–71.

Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge University Press.

McKendree, J., Stenning, K., Mayes, T., Lee, J., Cox, R. (1998) Why observing a dialogue may benefit learning. Journal of Computer Assisted Learning, **14**, pps. 110–119.

Monaghan, P. & Stenning, K. (1998). Effects of representational modality and thinking style on learning to solve reasoning problems. *20th annual meeting of the Cognitive Science Society*. Madison, Wisconsin. Pps. 716–721.

Stenning, K. Cox, R. & Oberlander, J. (1995). Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. *Language and Cognitive Processes*, **10 (3/4)**, pps. 333–354.

Stenning, K. & Yule, P. (1997). Image and language in human reasoning: a syllogistic illustration. *Cognitive Pscyhology*, **34**, pps. 109–159.

Appendix 1: video instructions transcription

Pre-test

> The purpose of this study is to let you know about a type of reasoning problem called a syllogism. First, you will be shown some examples of a syllogism, then after some practice with these, you will be shown a method of solving them. We are also interested in the effects of different methods of teaching. You will hear more about both topics when the course turns to graphical communication.
>
> So, what is a syllogism? Syllogisms are pairs of statements about different groups of people. Your task is to figure out what, if anything, follows necessarily from these statements. There are two examples of syllogisms on the first page of your booklet.
>
> Imagine that there are at least three groups of people at a party: students taking courses in architecture, in botany and in chemistry (As, Bs, and Cs). People at the

party may be taking no courses, a single course or combinations of courses. The information you get about the people at this party will be in one of four forms (these are called premisses):
You might be told that

- all As are Bs, or
- some As are Bs, or
- no As are Bs, or
- some As are not Bs.

The sorts of conclusion you can make will be in one of these forms, as well: either an all, a some, a no, or a some not. There may not be anything you can say for certain about the relation between the other two sets. In that case just write 'No conclusion'.

An important things to note:

- When you are told that some As are Bs, that means that at least some (and perhaps all) As are Bs.

Try answering the two practice examples on the first page of the booklet. If you have any questions, please ask the demonstrator now.

¡here the tape was paused, and students given an opportunity to ask the experimenter questions¿

There are 10 problems for you to answer. You have 10 minutes. Please do the problems in the order they appear in the booklet, and please don't go back to a previous problem. Try to distribute your time so that you attempt most problems.

POST-TEST

There is a method of solving these problems using diagrams. Some people like this method, other people don't. It doesn't matter if you find it easy or not, we are interested in how people differ in their approach to this method.

The three sets A, B, and C can be represented by three circles. Where circles overlap, this indicates that there are possibly people having both the properties indicated by the circles. To indicate where you know there are people, you put a cross in that part of the diagram. The different types of premisses can then be represented as follows (these are shown on the first sheet of the second booklet for reference) *¡see figure 1¿*

There is obviously more than one arrangement of circles which make the sentences true (in most cases). The arrangement shown is the one with the most regions representing the most possible sorts of thing.

Putting the premisses together means that the B circles have to be superimposed (they can get bigger or smaller as long as they keep the right relation to the other circle).

Here is an example of the syllogism *all As are Bs, some Bs are Cs ¡see figure 2¿*.

Once the B-circles are superimposed, the other two circles can take up different positions. We want to investigate what has to be the case in all cases, so we arrange A and C to make as many regions as we can, without changing the relation between each of them and the B-circle. In the example, the A- and the C-circle are made to overlap to make the most number of regions.

Now we look to see what has happened to the X-marked regions from the premiss diagrams. If they have been cut in two by the other circle, then the 'X' is taken away. This is because the 'X' could end up either side of the line, and so we can't tell for

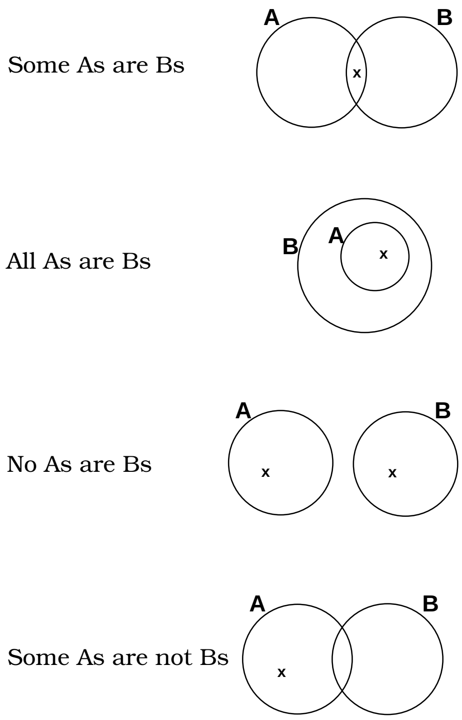


Figure 1: the premiss diagrams as presented on the second booklet

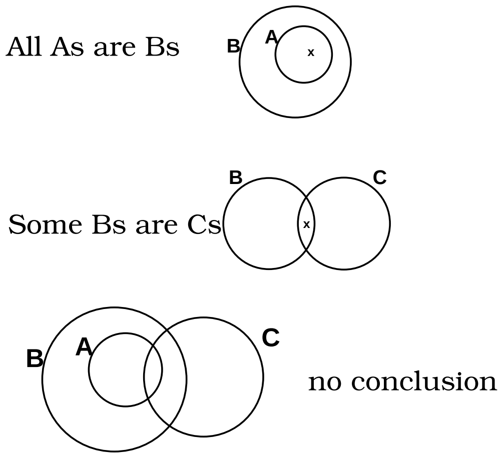


Figure 2: the example solution to the syllogism *aABsBC*

either half region, whether there must be anything in it. This has happened to both of the X-marked regions in the example, and so we don't keep either of the crosses.

Now, how do we use the diagrams to draw a conclusion? First, if there is no 'X' left, then there is no conclusion, because there is no type of person we are sure *must* exist. This is the case in the example, so there is no conclusion for this syllogism. If there is an 'X' left, then there is some region which corresponds to some type of person specified in terms of all three properties who must exist. We can work out a relation between A and C by describing this person. If the region with an 'X' in is an unbroken circle then the conclusion will be an all or a none, otherwise the conclusion will be a some.

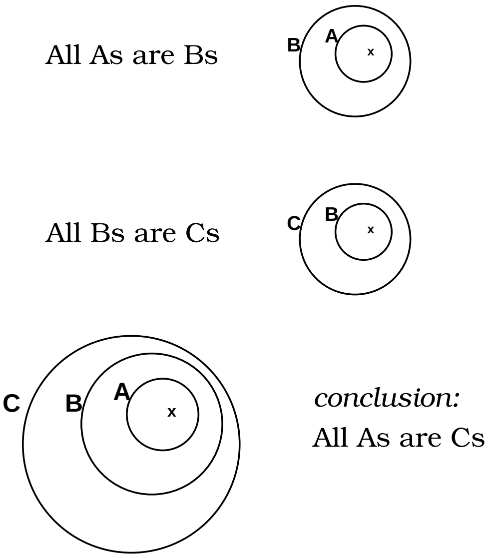


Figure 3: the example solution to the syllogism *aABaBC*

Here is another example of a syllogism solved using this method ¡*see figure 3*¿. This syllogism is *all As are Bs, all Bs are Cs*. Note that the X-marked region from the second premiss (from the B-circle) is removed because this region is divided up by the A-circle. There is one X left, though, and this means we can make a conclusion: *all As are Cs*.

¡*here the struggler group and the sailer group were shown video clips of students solving the syllogisms (see appendix 2)*¿

If you have any questions, ask the demonstrator now.

There are 40 problems in all, you have 20 minutes to answer as many of them as you can using this diagrammatic method. There are more problems than you are likely to get through, but try to work at the same pace as you did for the previous booklet. Please do the problems in the order they appear in the booklet, and please don't go back to a previous problem. You may start now.

Appendix 2: video clips transcriptions

Struggler

41.41
Instructor all architects are botanists

Student all As are Bs.

Instructor all botanists are chemists

Student all Bs are Cs. So all of that circle is in a big C circle.

Instructor And where do the x's go?

Student If all As are Bs then that's the x there, and if all Bs are Cs the x is there. All As are Bs ... But it could also go there as well, you could have two pluses but we've only got one there.

Instructor When we have all Bs are Cs if we'd drawn it separately, we would have seen that clearly I think. Just draw it now all Bs are Cs. And we have this X here. But we've got this region divided into two by the A circle, so we don't know where the x ends up, whether it's here or here.

Student so we cross that one out.

Instructor So we lose this x, but we keep this x. Why do we keep this x?

Student Because it hasn't been split up by the other circle.

43.36
Student Therefore all As are Cs.

44.05
Instructor all architects are botanists, some botanists are chemists.

Instructor Maybe draw it separately to begin with.

Student some Bs are Cs.

Instructor Then you put the B circles on top of each other but you maximise the number of possible crossings, the number of areas created.

Student that's the B circle, put that on top of it. And that's the A circle, do you have the C circle there?

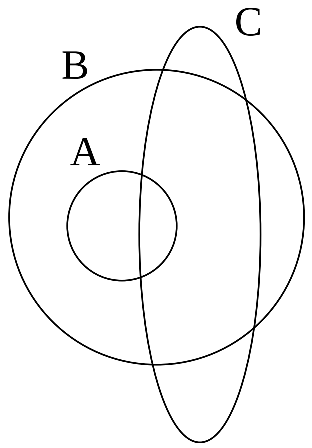


Figure 4: the "struggler's" diagram for the syllogism aABsBC

Instructor Yes.

Student but take that cross out which was originally there. I'm just getting a bit confused, A, B, and C. So the conclusion is...

Instructor What about this cross?

Student Oh, right. Does that disappear as well?

Instructor Why?

Student because we're crossing over the A section.

Instructor Right. So where is this region in the diagram?

Student That region is there.

Instructor OK, and it's been cut here, and so we don't know where the x ends up, whether it's in here or in here, so we lose it.

46.15
Student so the conclusion between As and Cs ... is it As and Cs? is that some C are A.

Instructor Why?

Student because that C and A overlap.

Instructor But we need to know for certain there's something in there then.

Student Oh but we don't. So that means not all ... (46.44) or no conclusion.

SAILER

40.46
Instructor all As are Bs, all Bs are Cs

Student so, I'm going to combine the two diagrams.

41.55
Student So that's that all architects are chemists.

43.33
Instructor all As are Bs, some Bs are Cs

Student mmm, well I've crossed that circle so I'll have to get rid of that cross because I don't know for definite which area of the new diagram it'd be in. And the same with this because the way I've drawn it I don't know for definite where the ... if a botanist is a chemist if he is also an architect. I do know that every A is a B that doesn't necessarily mean that he's also a C. Mmm, I seem to have got rid of all my crosses so I don't know.

Instructor If you don't have any crosses left that means you don't know where anyone is for sure, which would mean there'd be no conclusion.