Hats*

LaTeX file: *hats* — Daniel A. Graham <daniel.graham@duke.edu>, June 16, 2005

Imagine a class with n very logical girls sitting in a circle. The teacher enters the room with a bag of hats, informs the students that each hat in the bag is either white or red and then, walking behind the girls, places a hat on each girl's head. No girl can see the color of her own hat but each can see the color of every other girl's hat. We suppose, in fact, that all the hats are red.

Now the teacher states that she will give a prize to any student who can correctly identify the color of her own hat and a large enough penalty to any girl who specifies the wrong color to discourage guessing. The teacher then progresses repeatedly around the circle, asking each girl in turn whether she would like to state the color of her own hat or pass. Will any girl ever be able to state the color of her own hat? No, as we shall see, each girl will forever pass.

Now suppose the teacher announces to the class that at least one of the hats is red and then once again progresses repeatedly around the circle asking each girl in turn whether she would now like to state the color of her own hat. Will any girl ever be able state the color of her own hat? Yes, *every* girl will, in fact, eventually be able to state the color of her own hat. How the announcement of a fact that was already obvious to every girl produces such a dramatic change in the outcome is the subject to which we now turn.

To facilitate a graphical exposition, we consider the case in which there are three girls, let 0 denote white and let 1 denote red and use the binary number $c_1c_2c_3$ to denote the state in which c_i is the color of the hat worn by the ith girl. Thus 001, for example, denotes the state in which the first two girls are wearing white hats and the third girl is wearing a red hat. Since we are supposing that all the hats are red, 111 is actually the true state but there are eight possible states and these correspond to the vertices of the unit cube illustrated in Figure 1.

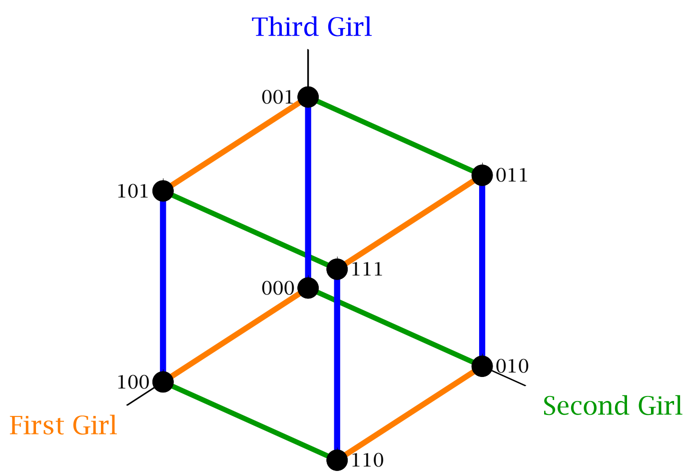

Figure 1: Interim Knowledge

Figure 1 also illustrates the private information partitions of the three girls at the "interim" stage in which each has had a chance to notice the colors of the hats of the other girls but the teacher has not yet made the announcement that at least one of the hats is red. Since the first girl cannot see the color of her own hat, her information partition is

$$\text{First} = \{\{000, 100\}, \{010, 110\}, \{001, 101\}, \{011, 111\}\}$$

*Based on the discussion in John Geanakoplos, "Common Knowledge", *Journal of Economic Perspectives*, 6 No. 4 (Fall), 1992, 53–82.

The pair of vertices within each of these possibility sets are connected in Figure 1 by a line *parallel* to the axis of the first girl. If you're lucky enough to be seeing this in color, these four lines are colored orange. Similarly, the information partition for the second girl is illustrated by the four green lines which are parallel to the axis of the second girl. Finally, the partition for the third girl is illustrated by the four blue lines which are parallel to her axis.

The common knowledge partition at the interim stage is perfectly course, i.e., it consists of a single possibility set containing all eight states. This follows from the fact that any state can be reached from any other state by a path from state to state that never leaves a possibility set for one girl without simultaneously entering a possibility set for another girl. At 111, for example, the first girl can think that the state is 011 and thus that the second girl can think that the state is 001. But this means that the first girl can think that the second girl can think that the third girl can think that the state is 000. Note that reaching 000 from 111 in this way involves moving around the top of the cube from 111 to 011 and then from 011 to 001 and finally down the back edge from 001 to 000. Since any state can be reached from any other in this way, common knowledge at any state include *every* state.

Now we suppose that the teacher will make one of two possible announcements, that there is at least one red had or that there are no red hats, whichever is correct. Since we are supposing that there are three red hats, the teacher will make the appropriate announcement with the effect illustrated in Figure 2. By announcing that 000 is not the true state, the connections to this state in each girl's information partition are broken. Alternatively, had the true state actually been 000 the teacher would have announced this and each girl's connections from 000 to other states would have been broken in the same way. The common knowledge partition after the announcement thus has two cells corresponding to the two "connected" components of Figure 2.

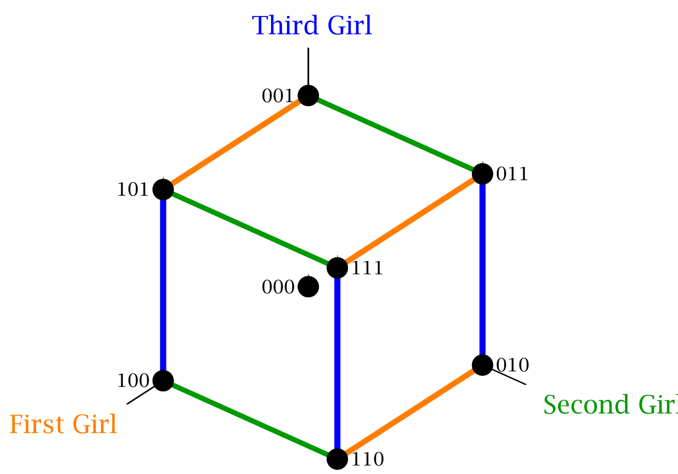

Figure 2: Teacher Makes Announcement

The story continues when the first girl is asked to state her own color. Since we are supposing that all three hats are red, the first girl will only know that the true state belongs to $\{011, 111\}$ and will therefore pass. Note that by passing, the first girl effectively announces that the true state cannot be 100, since she would have been able to deduce her color had that been the state and would not have passed. The effect of her announcement is to sever all ties to 100 and is illustrated in Figure 3. Note that the common knowledge partition now has three connected components.

When the second girl is asked to state her color she too will pass since she only knows that the true state belongs to $\{101, 111\}$. By passing, though, she effectively announces that the true state cannot be either 110 or 010 since she would have been able to deduce her color in either of these states. The effect of this announcement is illustrated in Figure 4. Note that the common knowledge partition now has four connected components and that it is now common knowledge that the third girl's hat is red.

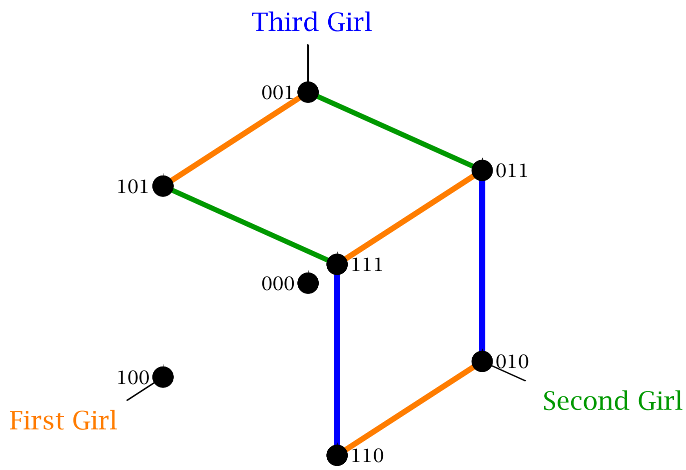

Figure 3: First Girl Passes

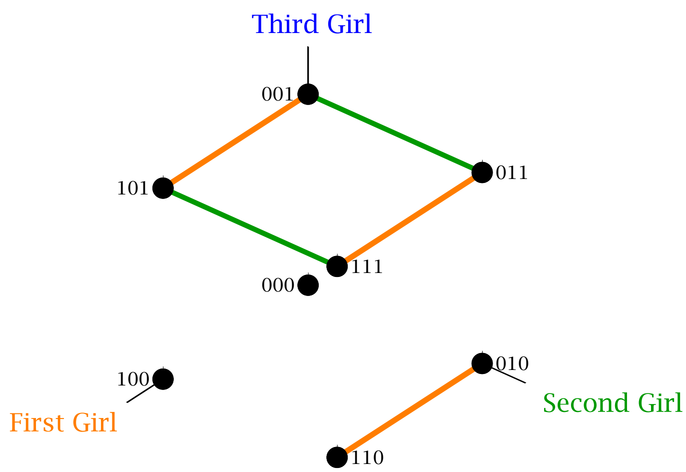

Figure 4: Second Girl Passes

When the third girl is asked to state her color, she will say “red” since her possibility set now contains the single state 111. Since the fact that the third girl’s hat is read was already common knowledge, this event conveys no new information. When the first girl is asked again to state her color, she will once again pass since she still only knows that the true state is in $\{011, 111\}$. The fact that she will pass is, in fact common knowledge after the third girl identifies her color since the first girl must pass at every state in the common knowledge possibility set corresponding to the top of the cube. And, since the fact that she will pass is already common knowledge, the fact that she does so conveys no new information. When asked, the second girl will also pass and this too was already common knowledge and thus conveys no new information. It is, in fact, common knowledge that the first two girls will continue to pass no matter how many times they are asked.

How would this story play out if there were no red hats? With only 1 red hat worn by the second girl? With 2 red hats worn by the first and second girls?

What would happen if the girls were required simultaneously to submit written statements about the color of their own hat or the word “pass” and the teacher then reported all of the statements and then repeated the process?