Experiment scenarios

Pedagogical conditions

Target blocks

On the next pages I've proposed two sets of target blocks.

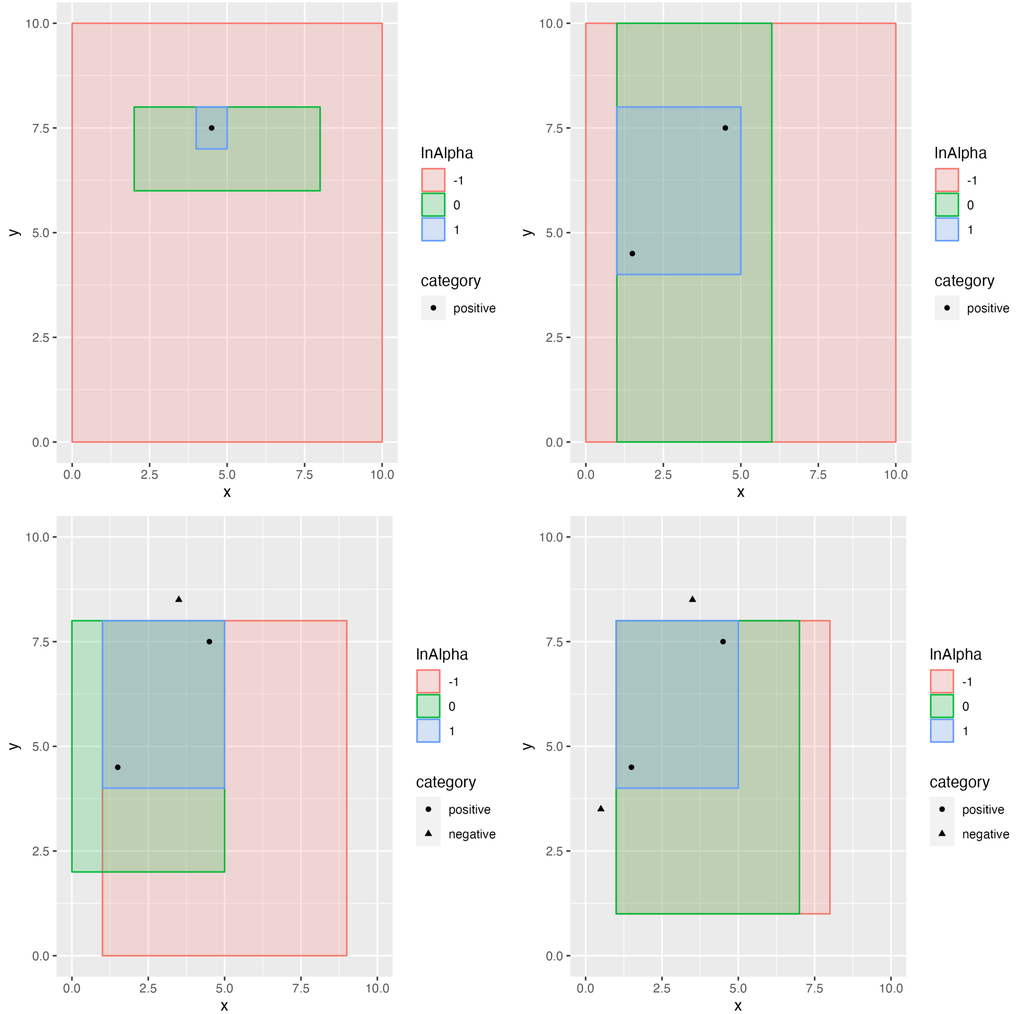
The plots show different predictions of learners, to demonstrate that these trials can distinguish between different learners.

Importantly, the two sets of blocks are actually functionally equivalent — the second block is just the inverse of the first block, which I think is useful because it keeps them controlled. Technically, people should respond exactly the same to both blocks, but I don't think they'll be able to tell the difference

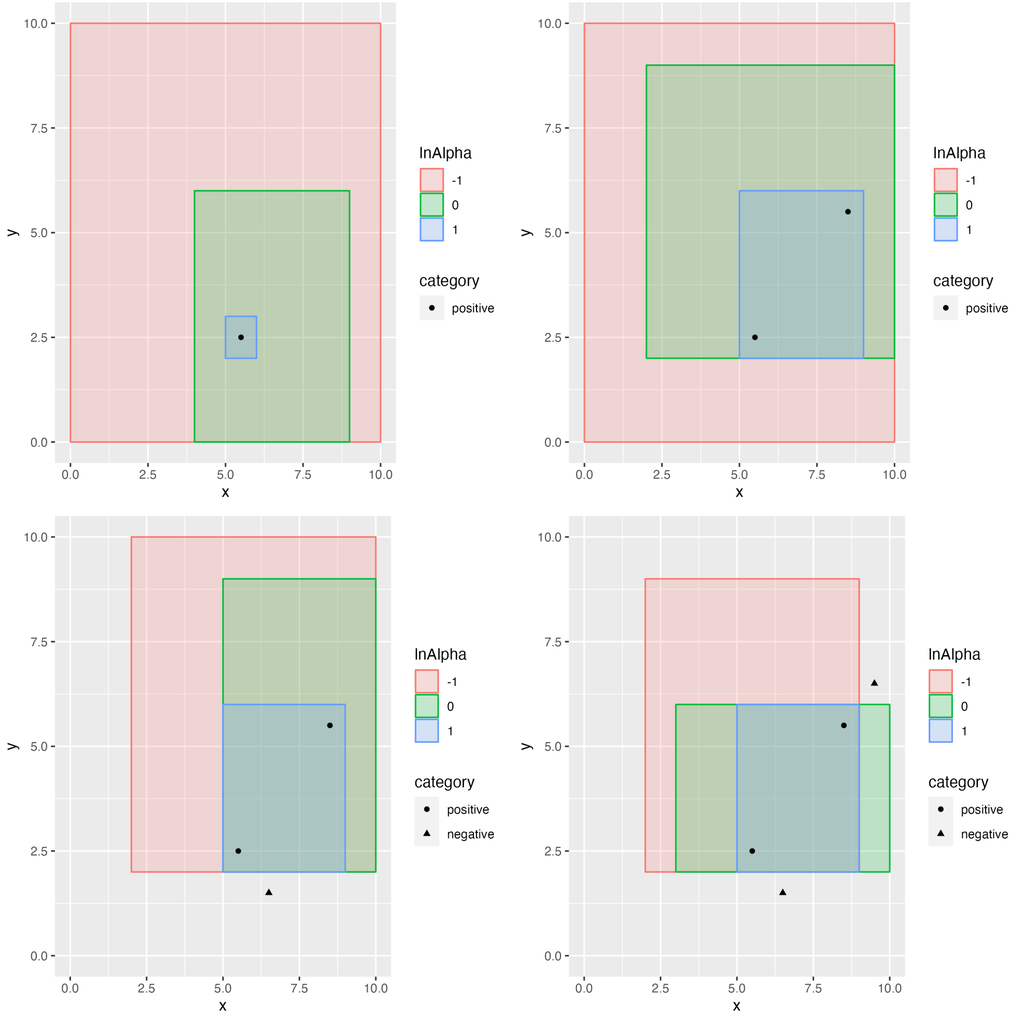
Target blocks

Blocks where the points are the same across all conditions

Target block 1



Target block 2



Filler trials

These trials are designed to actually look like what should be generated by a teacher from each condition.

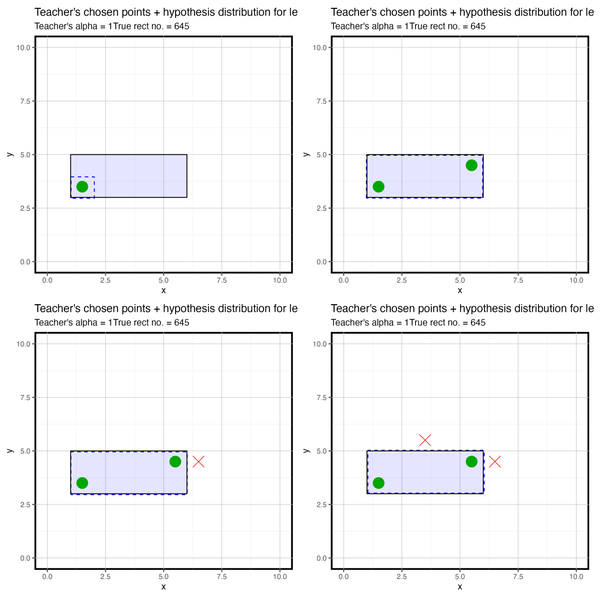
To choose them, I generated a bunch of predictions by the models for each condition (so they are actually generated by the correct model) and chosen which ones looked good.

Helpful teacher condition

Solid line = true rectangle, dotted line = learners prediction with alpha = 1

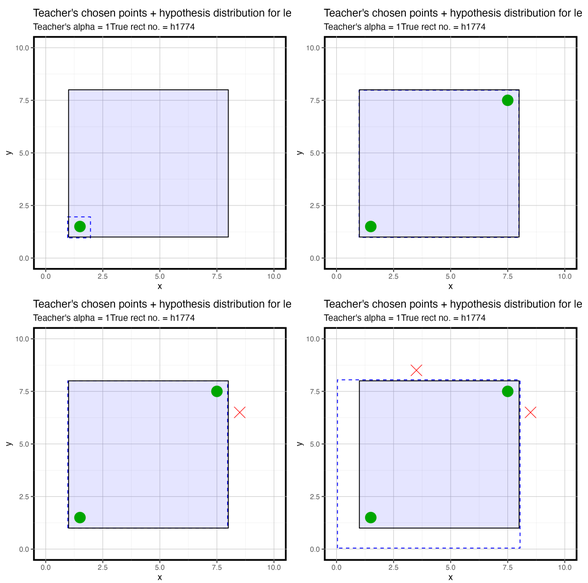
Teacher assumes learner is trusting

Block 1

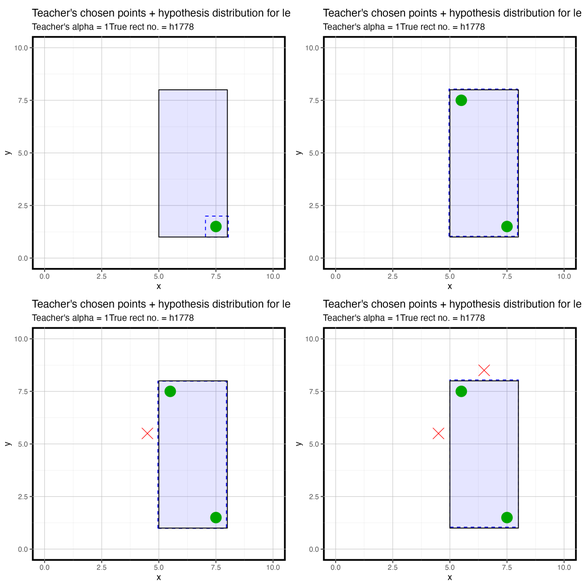


Block 2: target 1

Block 3



Block 4

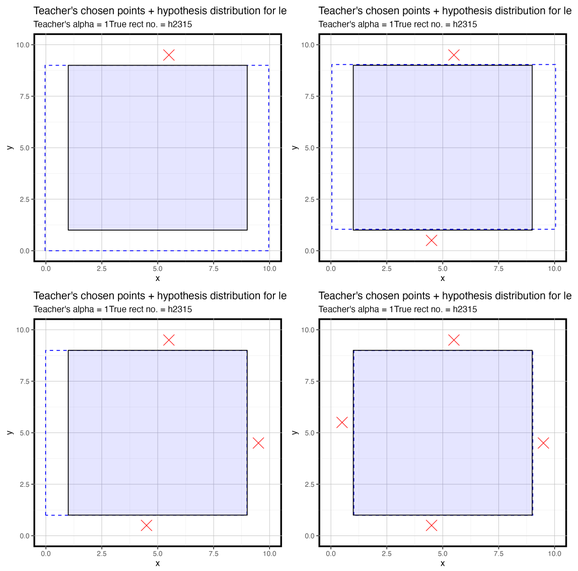


Helpful teacher condition

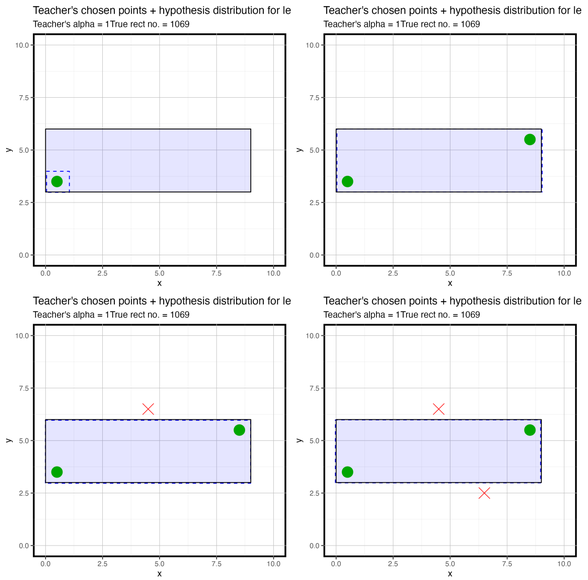
Solid line = true rectangle, dotted line = learners prediction with alpha = 1

Teacher assumes learner is trusting

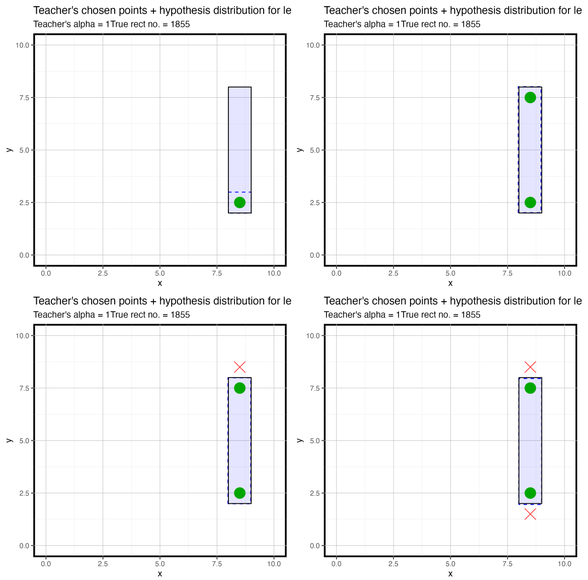
Block 5



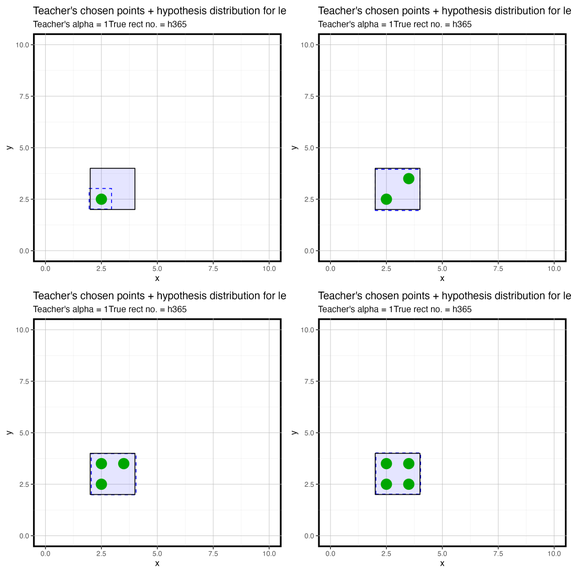
Block 6



Block 7



Block 8

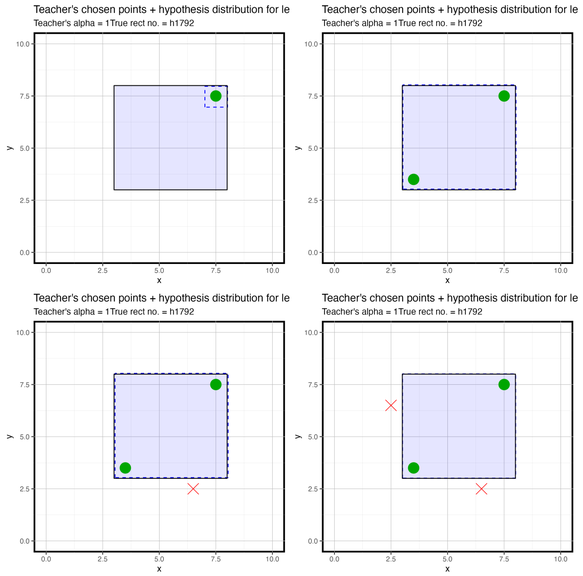


Helpful teacher condition

Solid line = true rectangle, dotted line = learners prediction with alpha = 1

Teacher assumes learner is trusting

Block 9



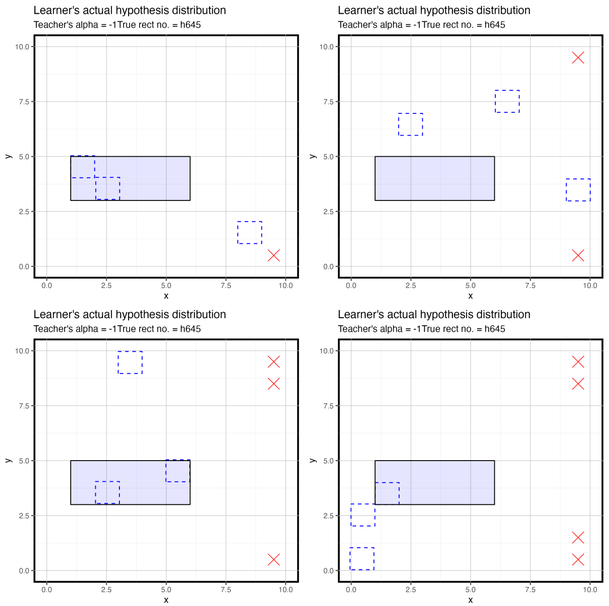
Block 10: target 2

Deceptive teacher condition

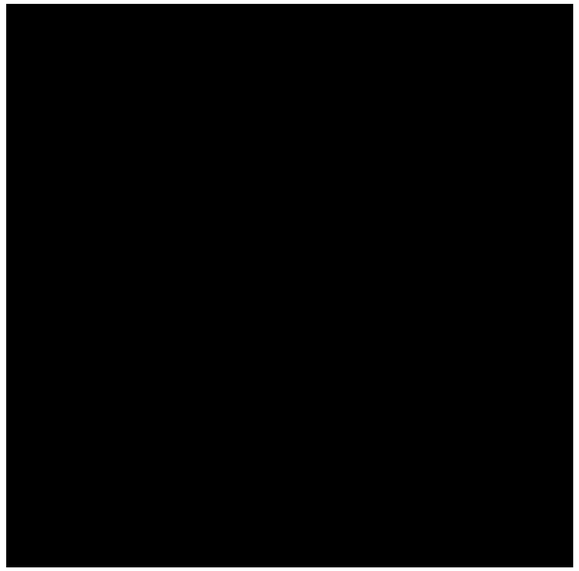
Solid line = true rectangle, dotted line = learners prediction with alpha = -1

Teacher assumes learner is trusting

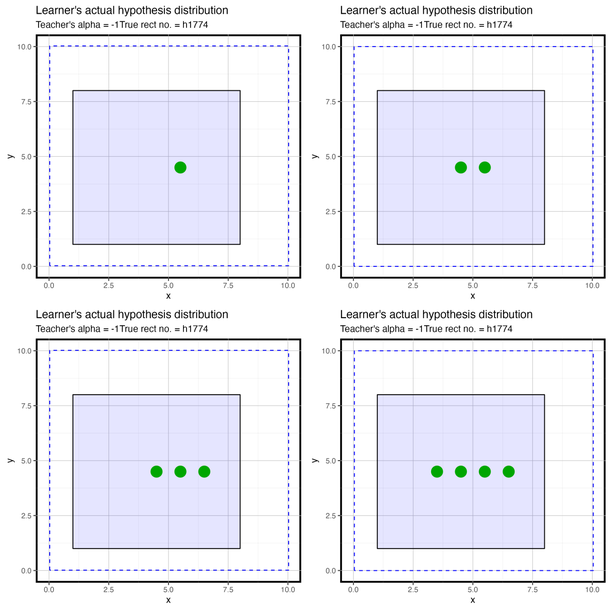
Block 1



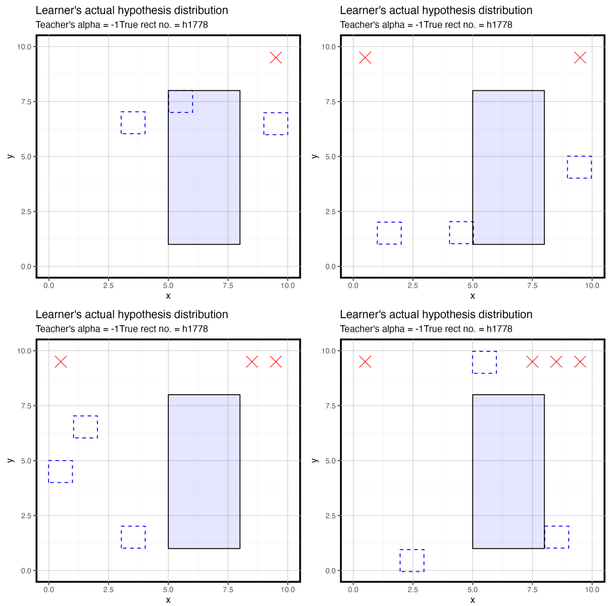
Block 2: target 1



Block 3



Block 4

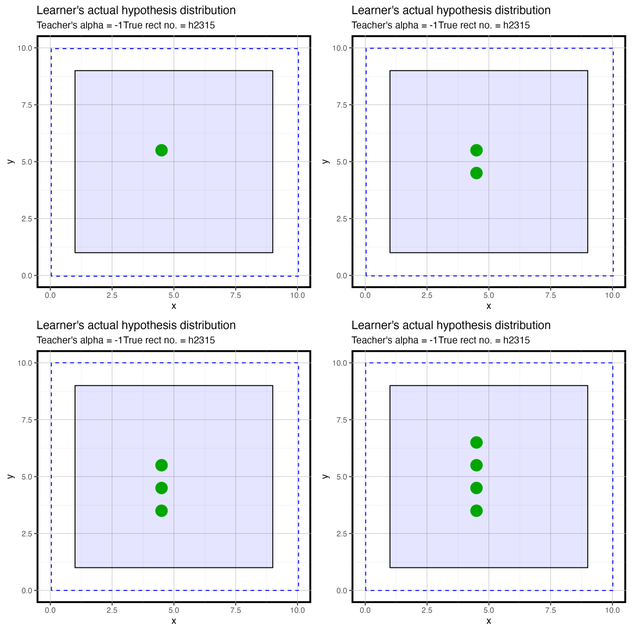


Deceptive teacher condition

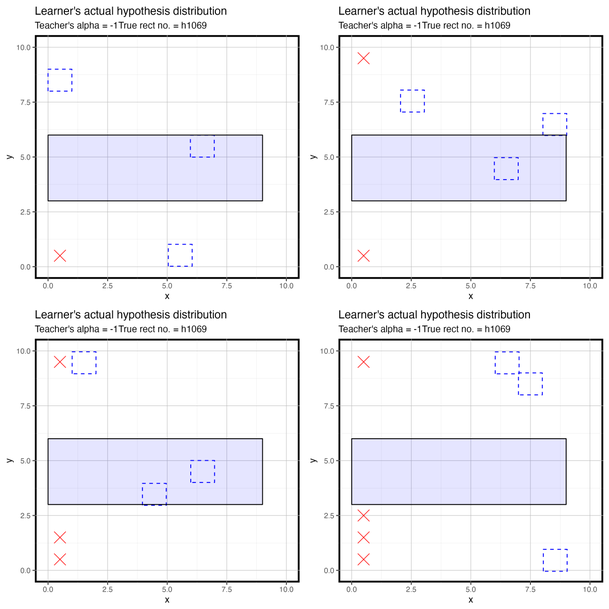
Solid line = true rectangle, dotted line = learners prediction with alpha = -1

Teacher assumes learner is trusting

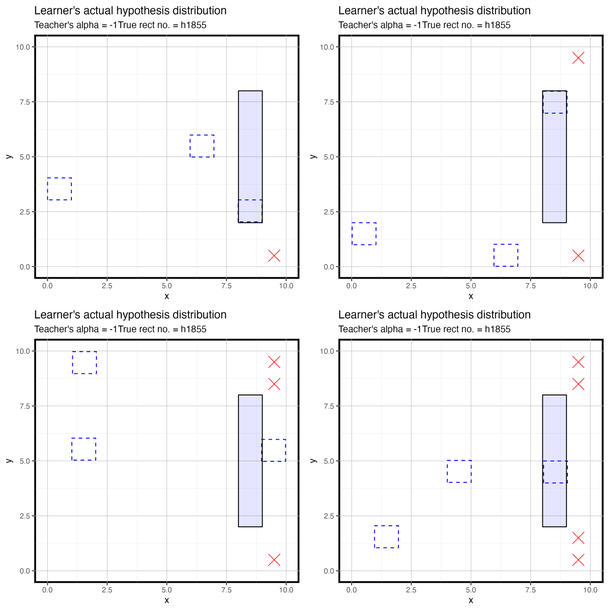
Block 5



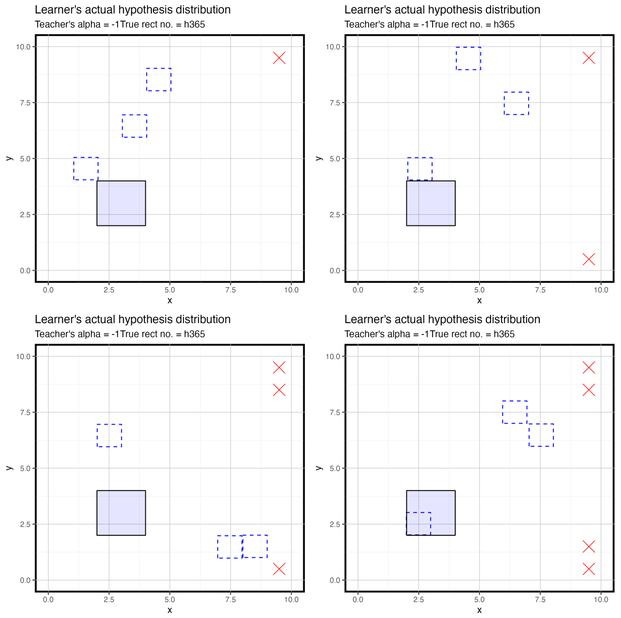
Block 6



Block 7



Block 8

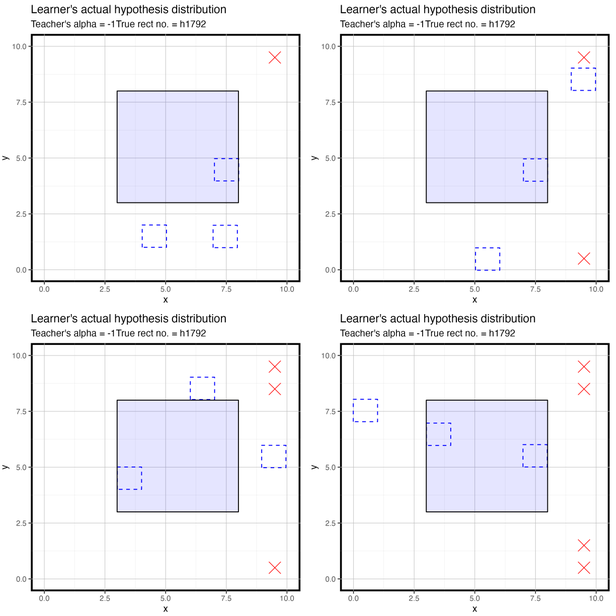


Deceptive teacher condition

Solid line = true rectangle, dotted line = learners prediction with alpha = -1

Teacher assumes learner is trusting

Block 9



Block 10: target 2

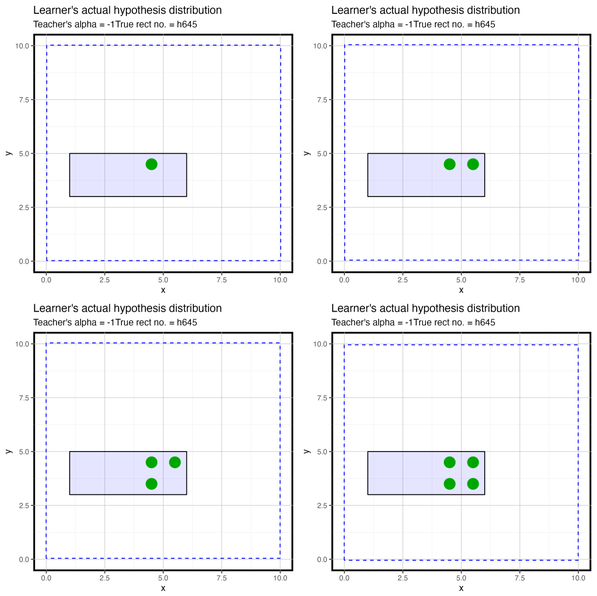
Some weird/repetitive patterns. What if the teacher assumes the learner is deceptive?

Deceptive teacher condition

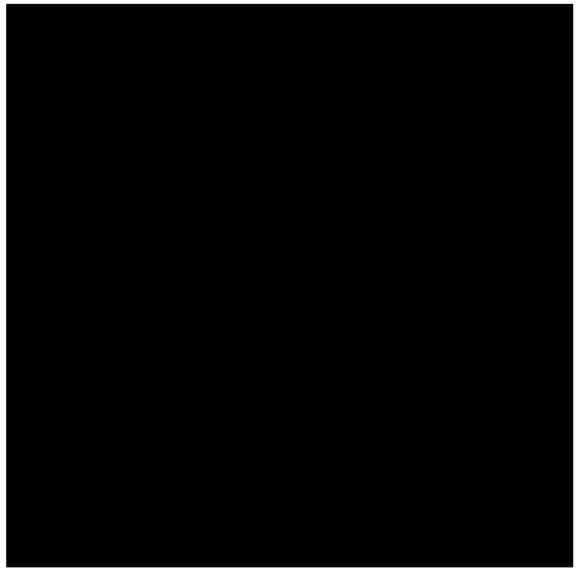
Solid line = true rectangle, dotted line = learners prediction with alpha = -1

Teacher assumes learner is mistrusting

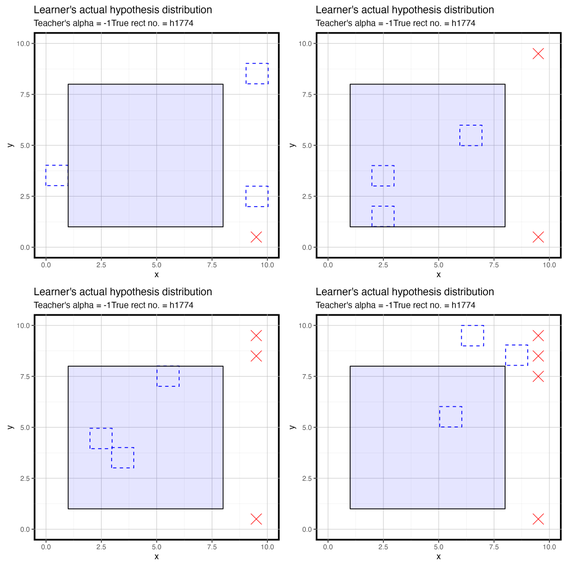
Block 1



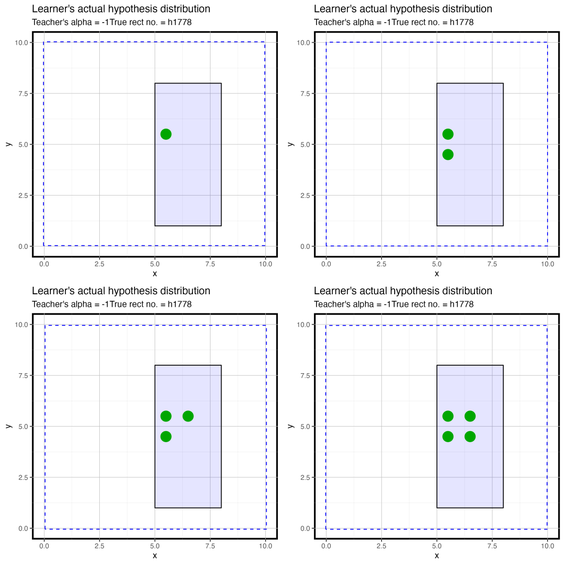
Block 2: target 1



Block 3



Block 4

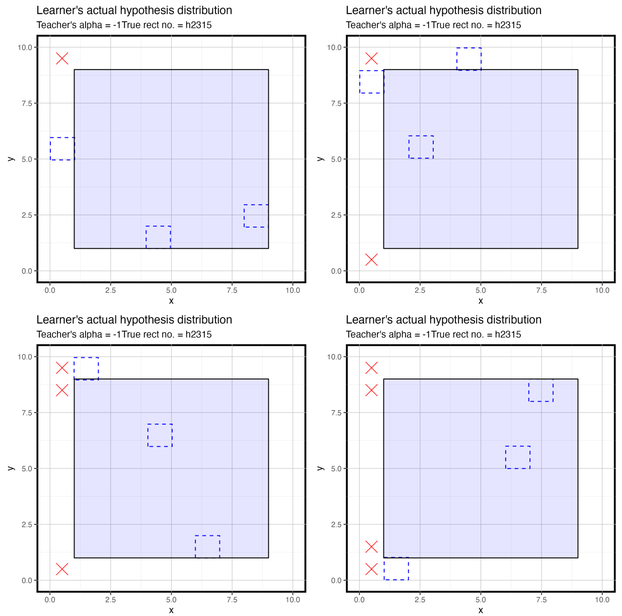


Deceptive teacher condition

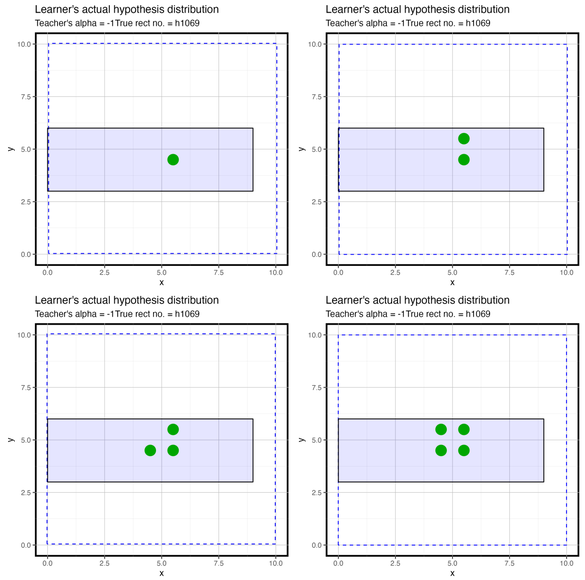
Solid line = true rectangle, dotted line = learners prediction with alpha = -1

Teacher assumes learner is mistrusting

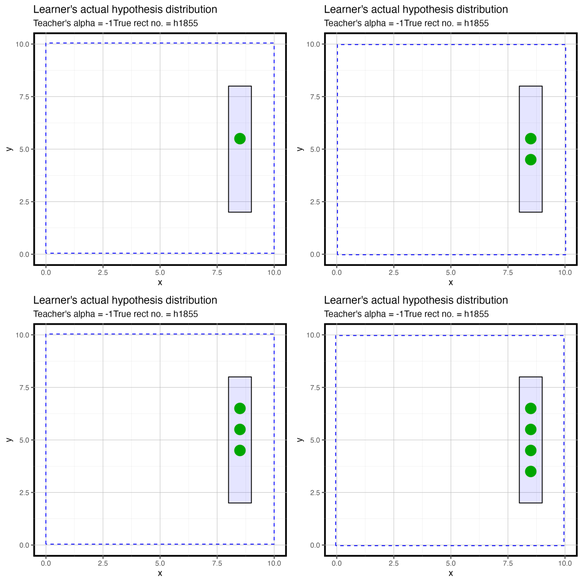
Block 5



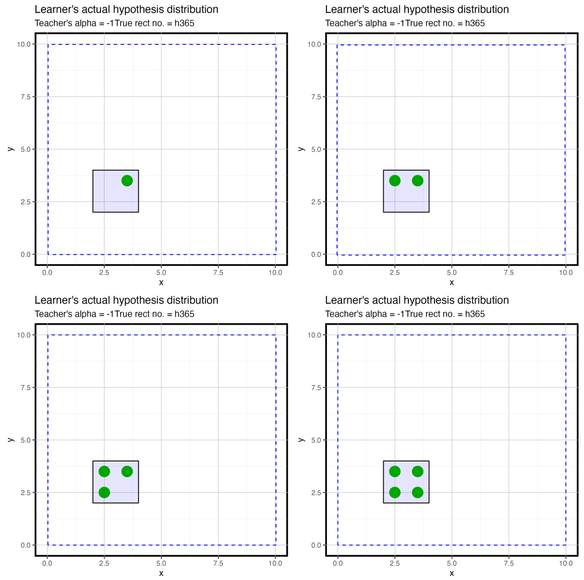
Block 6



Block 7



Block 8



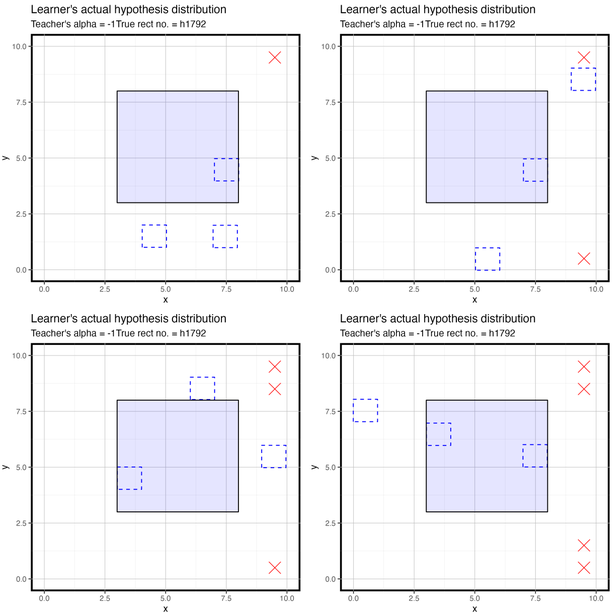
Blocks 7 and 8 are interesting to me — I guess since the teacher assumes the learner is mistrusting, even if they show them the right rectangle, they won't believe it?

Deceptive teacher condition

Solid line = true rectangle, dotted line = learners prediction with alpha = -1

Teacher assumes learner is mistrusting

Block 9



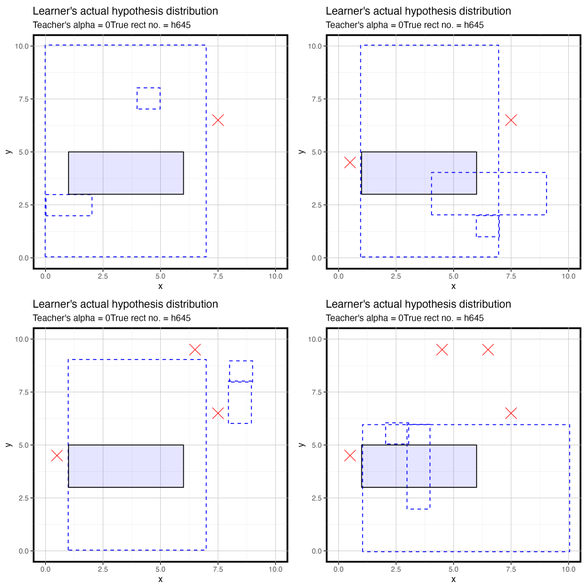
Block 10: target 2

Random (weak) teacher condition

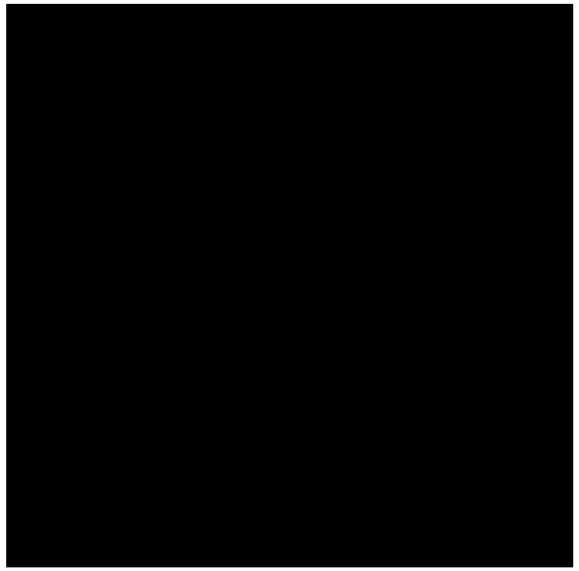
Solid line = true rectangle, dotted line = learners prediction with alpha = 0

Teacher assumes learner is trusting

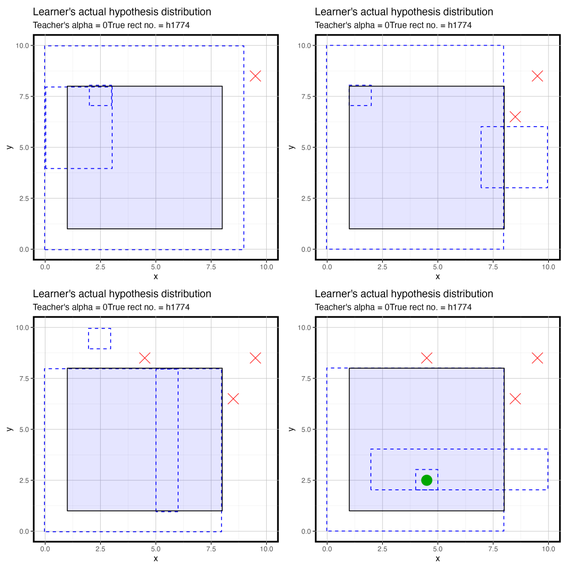
Block 1



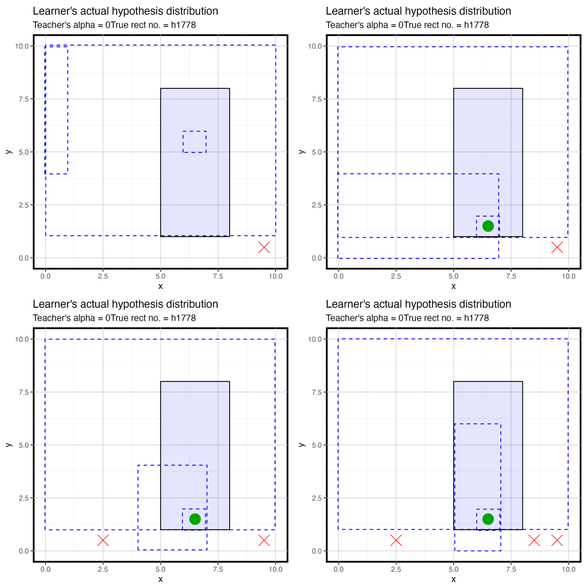
Block 2: target 1



Block 3



Block 4

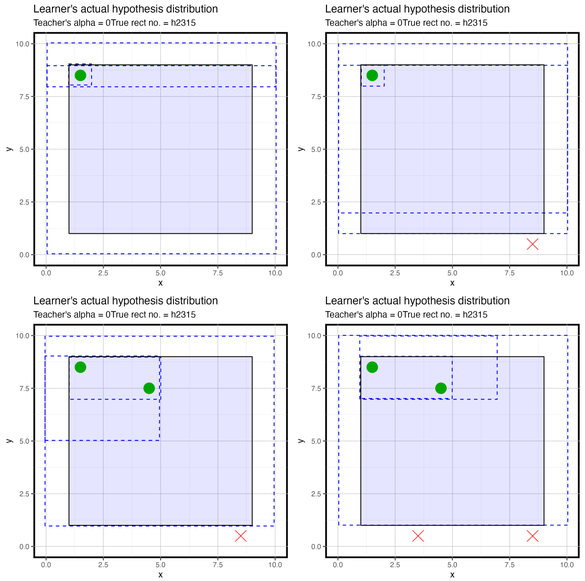


Random teacher condition

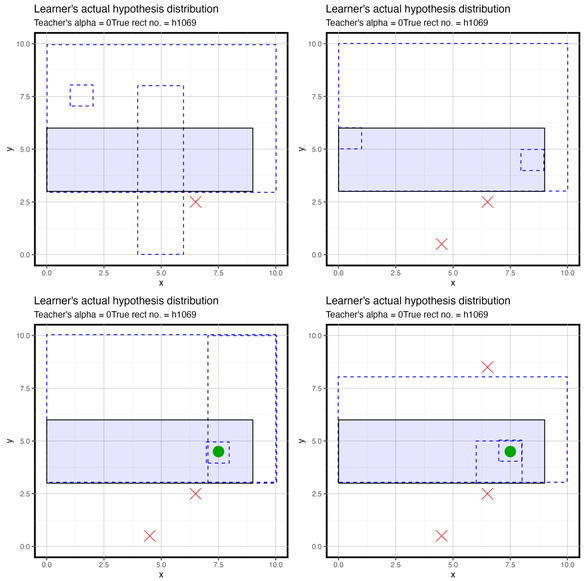
Solid line = true rectangle, dotted line = learners prediction with alpha = 0

Teacher assumes learner is trusting

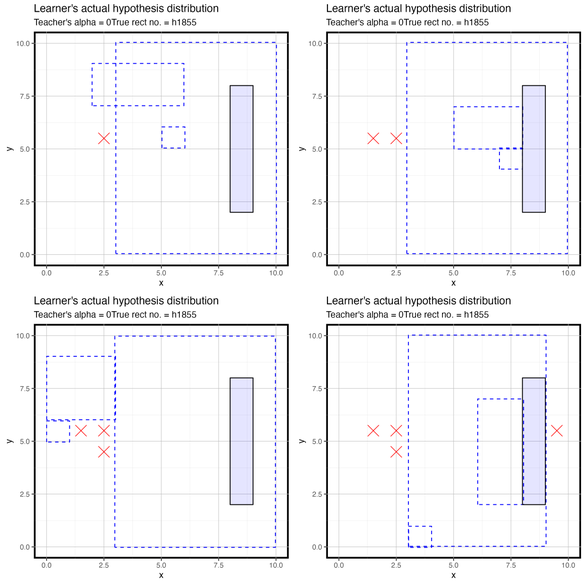
Block 5



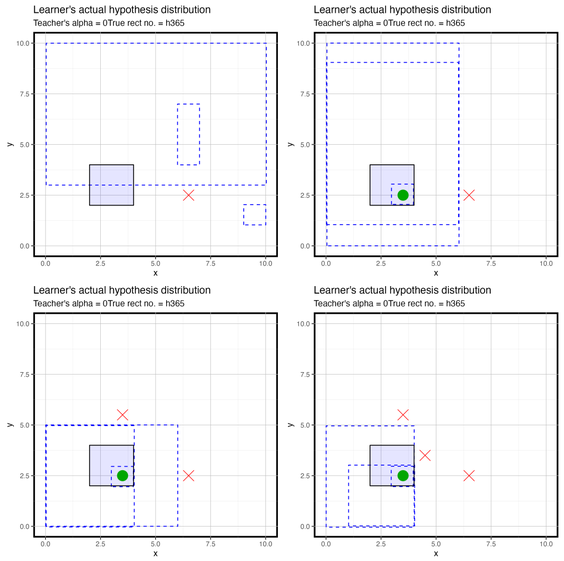
Block 6



Block 7



Block 8

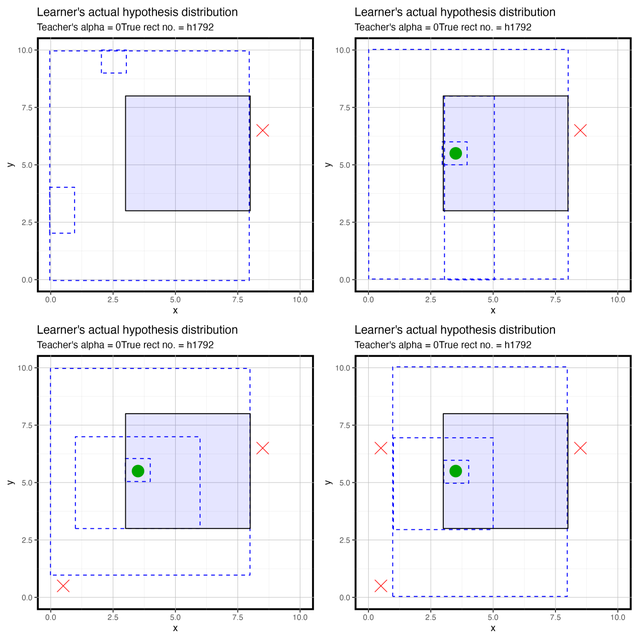
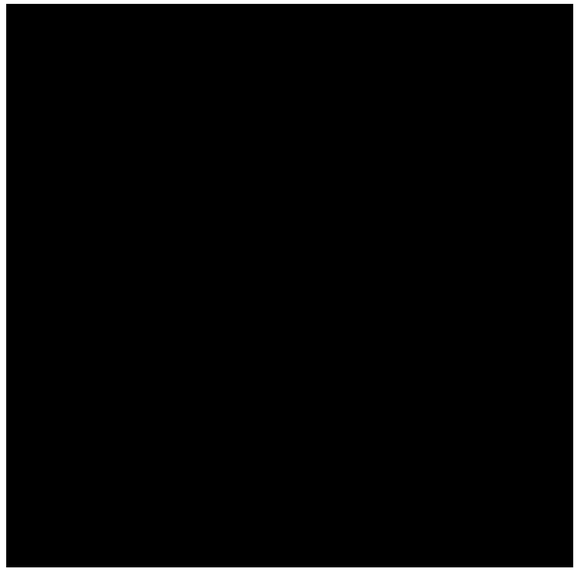


Random teacher condition

Solid line = true rectangle, dotted line = learners prediction with alpha = 0

Teacher assumes learner is trusting

Block 9



Block 10: target 2