

Unpacking responsibility

Nathalie Fernandez (nfern@mit.edu)

August 14, 2017

Contents

1 To do	2
2 Abstract	3
3 Research question	4
4 Experiment 1	4
4.1 Methods	4
4.1.1 Participants	4
4.1.2 Procedure	4
4.1.3 Design	12
4.2 Results	13
4.3 Discussion	14

1 To do

-

2 Abstract

When attributing responsibility to an agent for an outcome, we naturally use counterfactual reasoning. An agent is responsible if the outcome had not occurred but for their action. Think of an election where 5 people voted for a Republican and 0 voted for a Democrat. Assuming a majority rule, if someone, say Tom, would have voted for a Democrat, the outcome would not have changed. Yet we still attribute responsibility to the agent. Previous models have dealt with this by evaluating how close an agent was to being pivotal for the outcome. While these models can capture some scenarios, psychologists have found that when asked to consider an agent individually we tend to attribute more responsibility to that agent, and previous models fail to capture this. When considering an agent individually, we unpack their contribution: we tend to attribute more responsibility to that agent. To resolve this issue we modeled situations as hierarchical structures, where individuals form groups that contribute to an outcome. Representing groups as aggregate variables that could have acted differently, affects the pivotality of individual agents. For example, it's easier to imagine that Tom's vote would have made a difference, if two of the other voters are in a group that could have voted differently. We hypothesize that these hierarchical structures will affect how people make responsibility judgments, and we propose a model to capture these judgments. We have developed an experiment where we will compare participants' judgments to the predictions of our model.

3 Research question

4 Experiment 1

4.1 Methods

4.1.1 Participants

4.1.2 Procedure

Instructions

In this task you will look at employees' productivity.

You will have to determine how important each employee was
for the the overall productivity.

Continue

Figure 1: Instruction screenshot 1.

Instructions

The office has a goal to meet.



To evaluate whether the Office goal has been met, we must evaluate employee productivity.

[Go back](#)[Continue](#)

Figure 2: Instruction screenshot 2.

Instructions

Every employee has a goal to meet.

At the end of the week, they report their success or failure in meeting their goal.



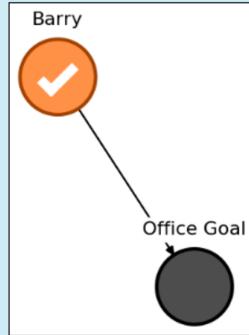
An X means failure. A check mark means success.

[Go back](#)[Continue](#)

Figure 3: Instruction screenshot 3.

Instructions

Some employees will contribute to the Office Goal.



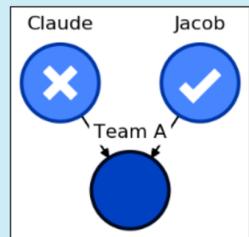
Here, the arrow means that Barry contributes to the Office Goal.

[Go back](#)[Continue](#)

Figure 4: Instruction screenshot 4.

Instructions

Others will contribute to teams. Teams and their employees share the same color.



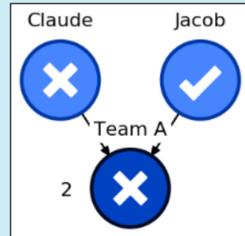
Claude and Jacob are in Team A.

[Go back](#)[Continue](#)

Figure 5: Instruction screenshot 5.

Instructions

In order for Team A to meet its goal, both Claude and Jacob have to succeed in their task.



The team failed because Claude did not succeed.

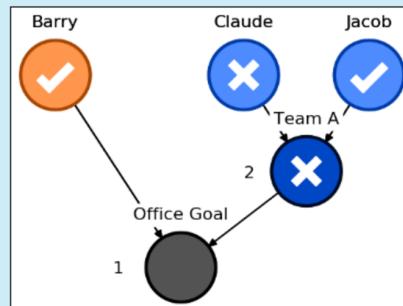
[Go back](#)

[Continue](#)

Figure 6: Instruction screenshot 6.

Instructions

Similar to employees, a team may contribute to the Office Goal.



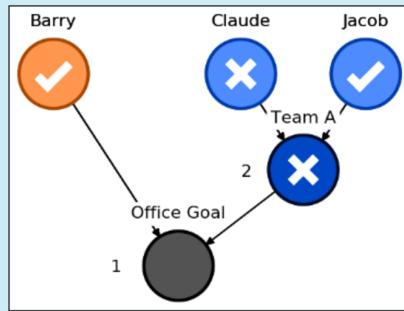
[Go back](#)

[Continue](#)

Figure 7: Instruction screenshot 7.

Instructions

The number 1 next to the Office Goal, indicates that at least 1 employee/team has to succeed in order to meet the Office Goal.



[Go back](#)

[Continue](#)

Figure 8: Instruction screenshot 8.

Instructions

A check mark means the office met their goal.



An X means the office did not meet their goal.



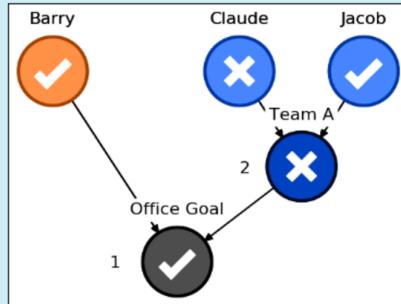
[Go back](#)

[Continue](#)

Figure 9: Instruction screenshot 9.

Instructions

To win the reward, Barry or Team A or both must succeed.



The Office Goal was met because Barry succeeded.

[Go back](#)[Continue](#)

Figure 10: Instruction screenshot 10.

Instructions

This concludes the introduction.

In the following page we will present you a set of questions to check your understanding of the task.

Then, you may begin the task.

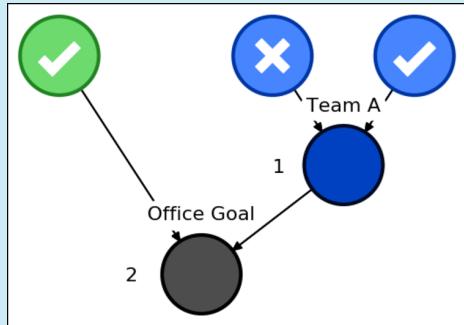
[Go back](#)[Continue](#)

Figure 11: Instruction screenshot 11.

Check questions

Please answer the following questions:

1. Based on the example below, evaluate the following.

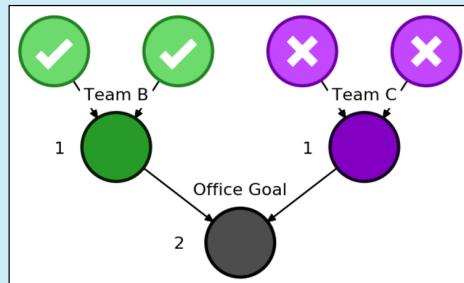


Team A: succeeded failed

Office Goal: was met wasn't met

Figure 12: Instruction screenshot 12.

2. Based on the example below, evaluate the following.



Team B: succeeded failed

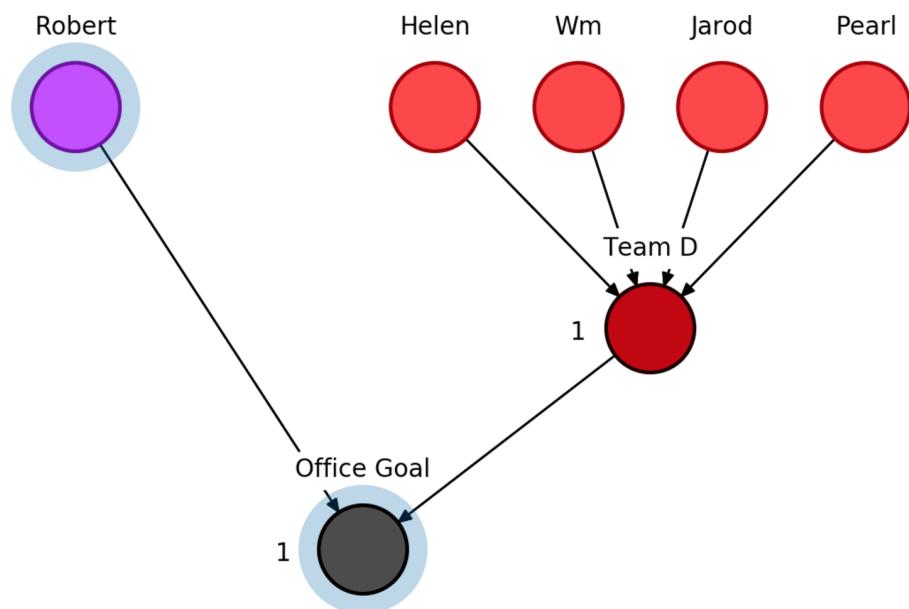
Team C: succeeded failed

Office Goal: was met wasn't met

Figure 13: Instruction screenshot 13.

Task

You will see several situations. To the left of the Teams and Office Goal, is the minimum number of employees or teams that need to succeed for the team to succeed or the Office Goal to be met.



How important is **Robert** for meeting the Office Goal?

not important at all

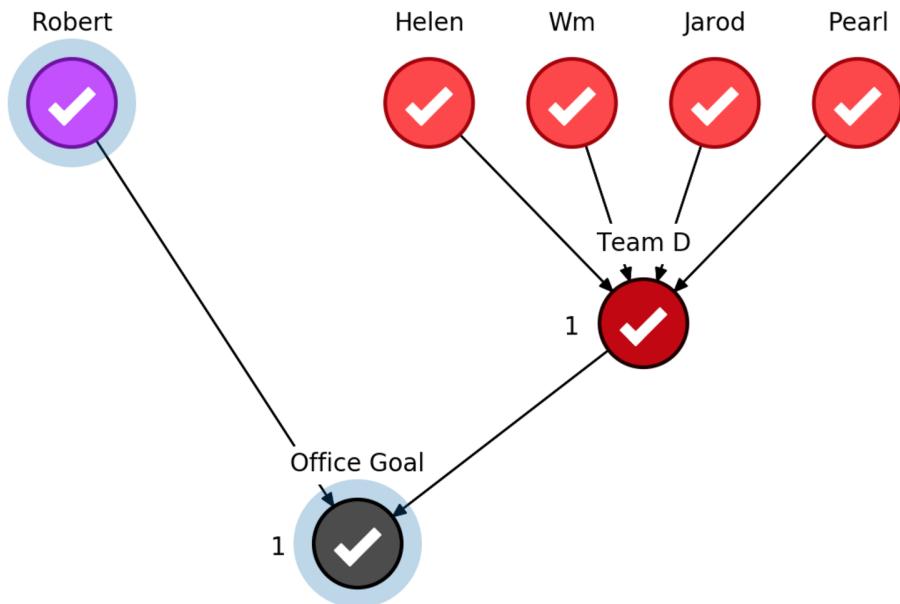
very important

[Next](#)

Figure 14: Instruction screenshot 14.

Task

You will see several situations. To the left of the Teams and Office Goal, is the minimum number of employees or teams that need to succeed for the team to succeed or the Office Goal to be met.



To what extent was **Robert** responsible for the Office Goal **being** met?

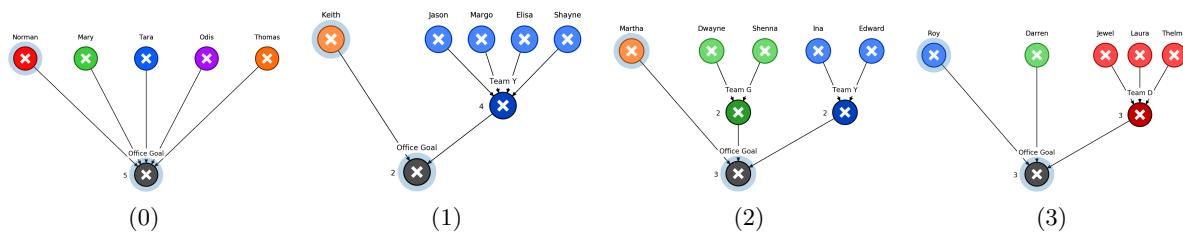
not at all

very much

Continue

Figure 15: Instruction screenshot 15.

4.1.3 Design



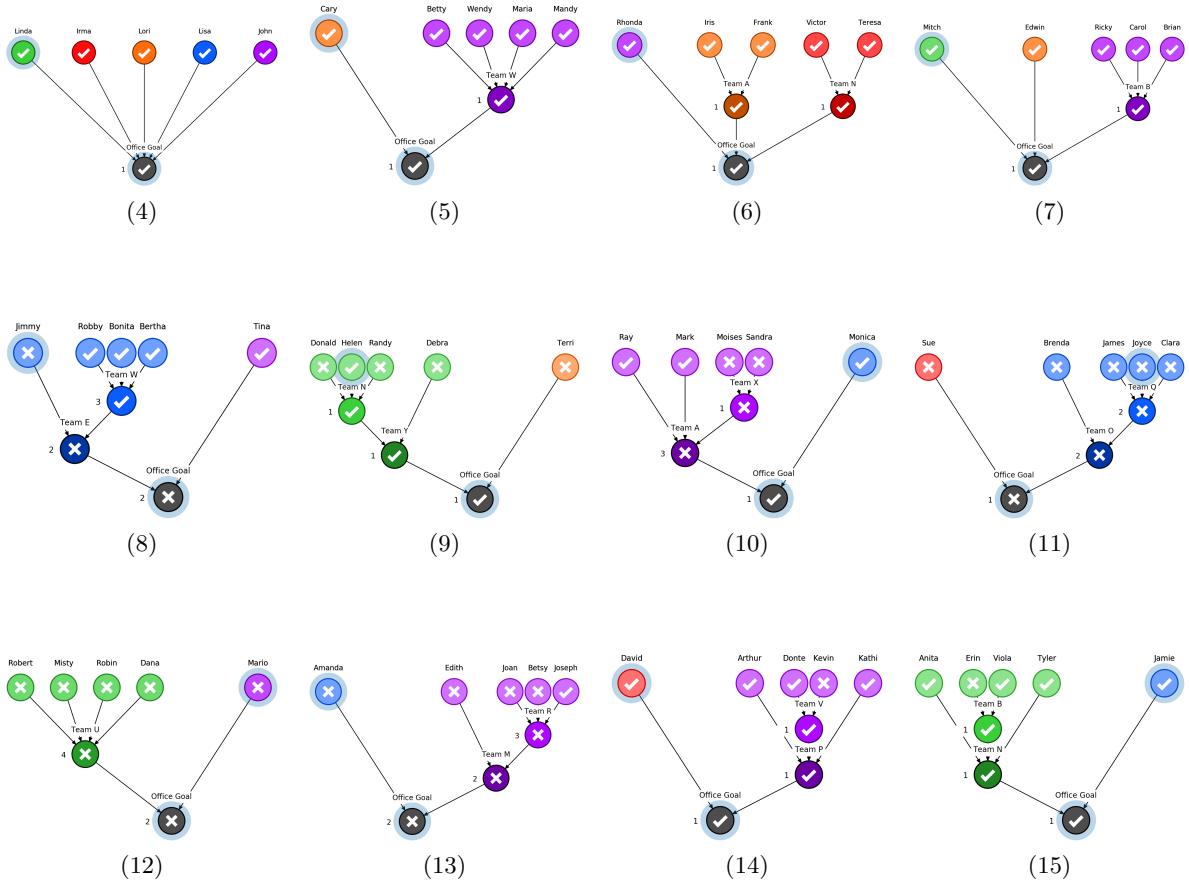


Figure 16: 16 Different situations used in Experiment 1.

4.2 Results

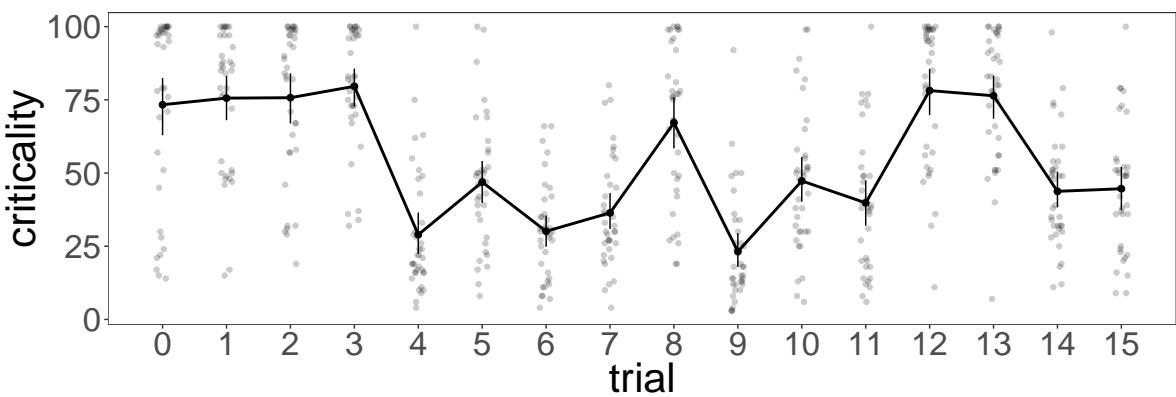


Figure 17: Criticality judgments: “How important is X for meeting the Office Goal?”

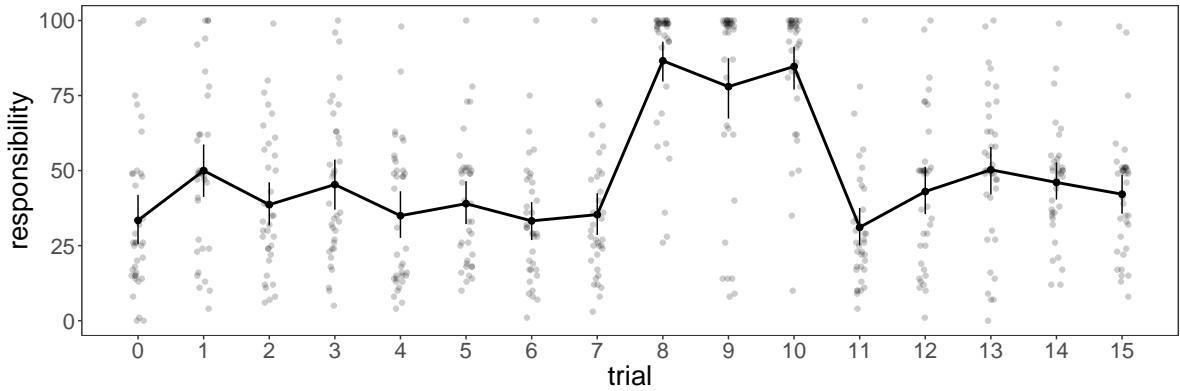


Figure 18: Responsibility judgments: “To what extent was X responsible for the Office Goal [not] being met?”

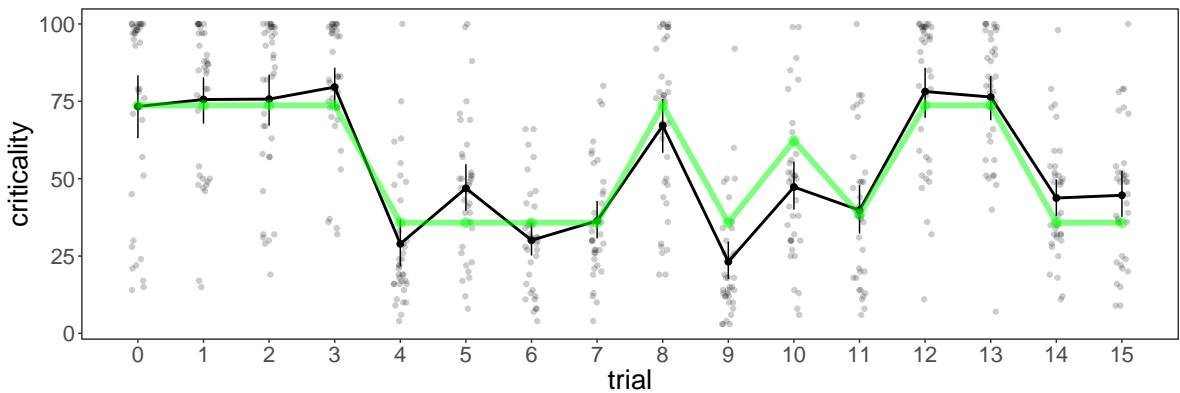


Figure 19: Criticality judgments: participant judgments vs. our model

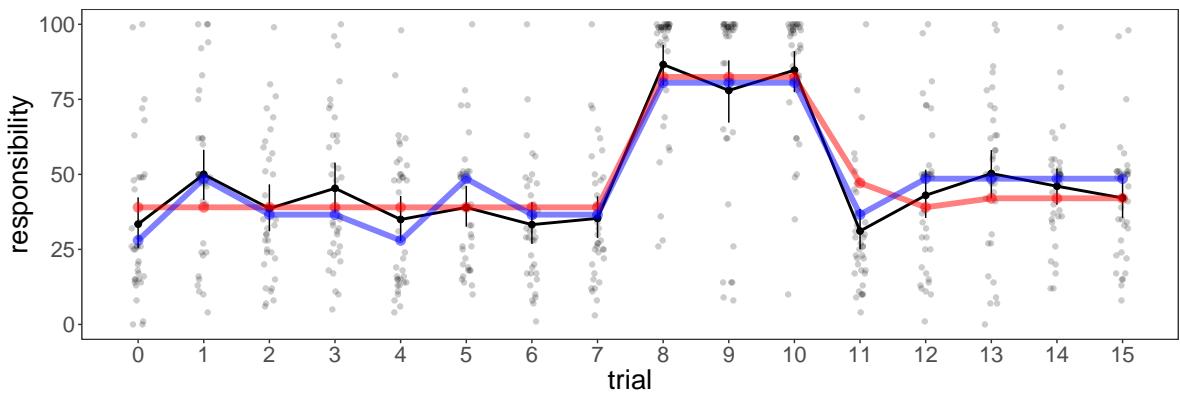


Figure 20: Responsibility judgments: participant judgments vs. our model

4.3 Discussion