1. Intro
   1. Cocaine addict in recovery
   2. Friend pulls out some coke and offers it to him
      1. Can imagine the internal debate
      2. Hand wants to reach for it, but he’s telling himself *don’t do it*
   3. Explain this w/ dual-system theory
      1. System 1 vs System 2
      2. Habitual, S-R associations say do it
      3. Goal-directed system says no
   4. Now imagine that addict just walks into a room where he used to snort
      1. No coke, but we know that still activates cravings
      2. He gets overwhelmed by cravings, drives to a dealer
   5. Hugely different scenario
      1. Before, the S-R habit made sense
         1. Saw a stimulus, his habitual system was compelling him to perform an action
         2. Executive, goal-directed system was saying no
      2. But now, it’s not like he habitually put on his shoes, coat, and drove to the dealer
         1. That’s all goal-directed behavior
         2. Seems like the goal-directed system was “hijacked”
         3. Habitual goal selection
   6. Possible objections
      1. What exactly do you mean by goal-directed vs habitual?
         1. I’ll clarify
         2. Doesn’t depend on vague systems division
      2. What’s really the difference between those two scenarios? i.e. whats the difference between a goal and a desire?
         1. A goal is a cognitive representation that coordinates behavior
         2. I’ve been thinking about it as.. Could a lizard do it?
   7. Automatic goal activation
      1. Tons of studies show that goals can be activated automatically by context/stimuli
      2. Goals implicitly associated with reward
      3. Also, those goals recruit executive resources
   8. But the authors of these studies have explicitly acknowledged that we don’t know how goals get automatically activated, or what the computational mechanisms behind these phenomena are
   9. Our goal
      1. Model it computationally
      2. Specify exactly the processes we’re talking about
      3. Use this computational specificity to demonstrate the phenomenon as simply and rigorously as possible
   10. Roadmap
       1. Describe formalization of the goals/habits distinction, and introduce our model
       2. Describe our behavioral experiments & results
       3. Discuss applications & future directions
       4. I’m gonna skip a lot of the computational specifics, i.e. the actual formulas we use
          1. I think the experiments are much more interesting
          2. If people want, I can go over the computational specifics at the end
2. Computational stuff
   1. Reinforcement learning
      1. Not sure how familiar everyone is with RL, so I’ll give a brief (oversimplifying) overview of the approach
      2. MDPs
         1. Can formalize the task facing an agent trying to choose behavior
         2. We have our agent (call him Bob)
         3. In an environment (maze)
         4. There are states (circles)
         5. Actions that can be taken from those states (arrows)
         6. And rewards/punishments (win or lose money)
      3. Agent’s goal is to maximize reward, minimize punishment
         1. How?
      4. Two types of solutions
         1. Builds model in head
            1. Forward planning
            2. Model-based
         2. Learns that red bad, blue good
            1. Model-free
         3. Captures goal-directed vs habitual
   2. How can you distinguish behavior that’s driven by model-based vs model-free learning?
      1. Give you a flavor for how the two algorithms operate
      2. Suppose we have a third option
         1. Can’t reach it on purpose
         2. Instead, each other option has a low-probability of transitioning to it
      3. Suppose person selects 1, but transitions to orange instead, and gets a big punishment
         1. What will each system say?
         2. Model-based system will say, I know that all options have an equally likely chance of going to 4, so that last punishment is irrelevant to my decision
         3. Model-free will say: 1 is bad!
         4. If people become less likely to choose 1, must be the habitual system’s influence
      4. For those of you who are familiar with the two-step task, this is different
         1. You’ll see why in a sec
   3. Our prediction
      1. In this scenario, in addition to learning that 1 is bad, the model-free system would also learn that setting a goal of blue is bad
      2. Why?
         1. In that trial, Bob set a goal of blue, and then got a punishment
         2. So it would learn that setting a goal of blue is bad
      3. How could we distinguish this model-free learning on *actions* vs on *goals*?
   4. Need to introduce two distinct actions which have the same goal
3. Our experiment
   1. Four numbers
      1. 1 and 3 both lead to blue
      2. 2 and 4 both lead to purple
      3. But distinct actions
   2. Bob is gonna play a bunch of trials
      1. On each trial, he’s given 2 of the 4 possible actions to choose from
      2. So like, 1 and 2
      3. He’s trying to choose the numbers which will ultimately give him the most points
   3. This looks easy, but the rewards are not static
      1. Instead, drifting over the entire game
      2. Blue could be good, but then could turn bad, and then purple might be good, etc.
   4. **Critically, rewards are tied to colors**
      1. So the two blue circles always give the same reward
      2. And the two purple circles always give the same reward
      3. So options 1 and 3 have the same goal: get a blue shape
   5. Critical subset of trials, which go like this
      1. Bob is presented with options 1 and 2
      2. He chooses 1, has an unlikely transition to orange
      3. Gets a big punishment (fade others)
   6. What will each system say?
      1. Model-based system says, each option has an equal probability of going to orange
         1. (We told them that explicitly, and they also learned it in practice trials)
         2. I’ve learned nothing
      2. Model-free action system has learned that 1 is bad
      3. Model-free goal system has learned that setting the goal of blue is bad
   7. But now suppose on the next trial, Bob has a choice between 3 and 4
      1. Model-based system will not incorporate the punishment into decision (fade)
      2. Model-free action system will not incorporate the punishment into decision (fade)
      3. But model-free goal system will!
         1. More likely to choose 4
      4. So if the value of orange affects people’s choices on the next round, then it can only be due to a model-free goal system
   8. Results
      1. We ran this on people
      2. Looked at all the trials on which people got rewarded versus when they got punished
         1. Looked at the proportion of trials on which they subsequently chose the same goal
         2. Significant difference
      3. We also ran a mixed effects model where orange’s value predicted people’s choices
         1. The effect was highly significant
         2. Even when controlling for MB and MF value of 3,4
      4. If anyone asks how we estimated the MB/MF value, tell them
      5. Whole procedure was validated w/ generative model
   9. Problem
      1. Could be accounted for by hierarchical model-free models
      2. i.e. there’s a higher-order “goal” state which the model-free system learns on, but is not really a goal in the true, flexible planning sense
      3. Wanted to show that this model-free learning was happening on true goals
4. Experiment 2
   1. Same setup as before
   2. Except we also trained people on a separate set of transitions
      1. Letters to numbers
   3. On most trials, the game was exactly like before
   4. Except for critical trials
      1. Choice between 1 and 2, choose 1, transition to orange, get punishment
      2. But THEN, choice between C and D
      3. If Joe has really learned that choosing blue is a bad goal, they should be less likely to choose C
   5. Results
      1. Same as before
   6. Problem
      1. You might think that, in their heads, people are just combining 1 and 3 into one “action” and 2 and 4 into one “action”
      2. Then it could really just be model-free action learning
5. Experiment 3
   1. Similar setup as before
   2. Except now, instead of all circles, we have different shapes also
      1. Two circles, two squares, and a triangle
   3. And two different types of trials: color trials and shape trials
      1. On color trials, what mattered was the *color* of the objects
         1. i.e. blue objects had a certain reward, purple objects had a certain reward, etc.
      2. On shape trials, what mattered was the *shape* of the objects
         1. Circles had a certain reward, squares had a certain reward, etc.
      3. Subjects are informed before each trial whether it’s a color or shape trial
   4. And also two types of critical trials
      1. On *congruent* critical trials…
         1. Let’s say it’s a color trial.
         2. Choice between 1 and 2, choose 1, transition to orange, get punishment.
         3. Next trial, *it’s also a color trial*. Choice between 2 and 3.
            1. Why 2 and 3 instead of 3 and 4?
            2. It will make sense in a second
            3. But note that the logic of the trial is exactly the same
         4. Person has learned that a goal of blue is bad, so less likely to choose 3.
      2. But on *incongruent* critical trials…
         1. Same thing. Color trial, choice between 1 and 2, choose 1, transition to orange, get punishment.
         2. But next trial, *it’s a shape trial*. (Note that we had to do 2 and 3 here instead of 3 and 4, because 3 and 4 both have the same shape).
         3. If person has just merely learned that 1 and 3 are the same action, they’ll be less likely to choose 3.
         4. But, if they’ve really learned that the goal of getting blue is bad, the value of orange will have *no impact* on their choice.
         5. Because on shape trials, you can’t set a goal of getting blue!
      3. So we predict that we’ll see our results only for *congruent* critical trials
   5. Results
      1. With data collapsed across participants, we see a significant difference for congruent critical trials but not incongruent critical trials
      2. We also estimated a mixed effects model w/ the interaction between orange’s values and the type of critical trial
         1. Coded incongruent trials as 0 and congruent trials as 1, so a positive interaction would mean that the effect was stronger for congruent trials
         2. Interaction was significant
      3. If anybody asks, no main effect of orange’s value (which means that the incongruent trials did not show a significant effect even in the mixed effects model)
   6. Roadmap
      1. Phew
6. Applications
   1. What’s new here?
      1. We already knew that goals can be activated automatically, and that they’re sensitive to reinforcement
         1. But nobody has formalized that w/ reinforcement learning, or modeled it so precisely
      2. A few computational models that hint at our results
         1. Algorithm in computer science called Dyna-2 that has MF values influencing MB choices
            1. But not meant to model human decision making.. They used it to play Go
         2. Pavlovian pruning of decision trees
            1. Showed that people reflexively stop considering decision branches if there’s a big loss in them
            2. Influence of a “less rational” system on goal-directed planning, but not the same system or integration mechanism as what we’re talking about
            3. Also, only destructive, not constructive
         3. MF gating of working memory
            1. Michael Frank has demonstrated that the dopaminergic learning mechanisms in the basal ganglia often associated w/ model-free learning gate information coming in & out of working memory
            2. System learns that holding certain pieces of information in working memory is linked to reward/punishment
            3. Similar idea as ours, except we’ve applied it to goal selection rather than working memory

Probably intimately related

Could view our work as an application of this neurobiological model to higher-level executive function

* + 1. But this is one of the first studies that rigorously demonstrates a model-free influence on model-based goal selection, or that provides a computational mechanism behind automatic goal selection
  1. What does this help explain?
     1. Properties of goals
        1. Automatic goal selection & reinforcement
        2. Goals can be intrusive
           1. People with frontal lesions automatically act on certain goals when presented with contextual stimuli
        3. Goal selection often gets “fixed”, like a habit
           1. Called “functional fixedness”
           2. Once people learn a certain amount of things you can do with an object, they tend to stick to that limited set
     2. Bigger picture stuff
        1. Acquiring complex cognitive skills
           1. Lots of research shows that training incrementally improves our ability to solve complex cognitive tasks – like long division
           2. How?
           3. Our model hints at an answer
           4. We know that people solve complex tasks by breaking them down into hierarchies of goals and subgoals

You might be doing a problem, and suddenly you come across a part where you need to do long division

Holding that superordinate goal in mind, you then delve into the 5 subgoals of long division, which each have their own subgoals

* + - * 1. One way that people might get better at these tasks w/ training is if the context of a superordinate goal, like “do long division”, can automatically activate the appropriate subordinate goals
        2. Habits of thought
      1. Addiction
         1. It’s been known for a long time that cravings can be activated by contextual cues
         2. This provides an explicit mechanism for what’s going on there
         3. Shows why *implementation intentions* work well… You need to activate a different goal in the face of those contexts
      2. Doctrine of double effect
         1. Idea that it’s not okay to harm someone as a means to an end, but it’s okay to harm someone as a *side effect* of another action

Used in justifying civilian “collateral damage” in bomb strikes

Also, euthanasia

People often think that killing somebody purposively is not okay

But if they die as a side effect of pain-reducing opiates, that’s okay

* + - * 1. Why? Person is equally dead either way
        2. To harm somebody as a means to an end, you have to mentally represent the goal to harm them

And every time you (or others) have had that goal in the past, you’ve been punished for it

If our model is correct, you will acquire an aversion to that goal

* + - * 1. On the other hand, if you harm somebody as a side effect, you don’t need to have the goal of harming them

So you won’t get the same aversive reaction

* 1. What’s next?
     1. fMRI version
        1. Trying to see if we can detect a goal prediction error signal
     2. More realistic version
        1. For example, instead of choosing different colored objects, we could have people be choosing between buying different stocks/bonds
        2. See if a random loss on one stock negatively affects people’s goal of buying stock