



## PRECOMMITMENT: A WAY AROUND TEMPTATION

Zeb Kurth-Nelson<sup>1,2</sup> and A. David Redish<sup>3\*</sup>

<sup>1</sup>University College London, London, United Kingdom, <sup>2</sup>DeepMind, London, United Kingdom, <sup>3</sup>University of Minnesota, Minneapolis MN, United States

### REVIEWED BY:



**GABI**  
11 YEARS OLD

### IMPULSIVITY

Wanting to get rewards immediately, even though it might be better to wait.

Impulsivity means we want things that make us feel good now, and we want to postpone things that take effort. This means that what we want changes—today we want to do the more difficult thing tomorrow, but tomorrow, we will want to take the easy option and will no longer want to do the thing that takes effort. One potential solution to impulsivity is something called precommitment, where we set things up today to remove the easy option tomorrow. It turns out that precommitment depends on the fact that the brain contains lots of different opinions—like a group of friends deciding what to do. Learning how to help yourself precommit can help you to avoid temptation and complete those difficult tasks that need to get done.

### NOW OR LATER?

Sometimes we need to make decisions between things we want now and things that will be better for us in the long run. Imagine it's Thursday morning and there's a big test on Friday. Your coursework and a new video game are both sitting on your desk. What would you rather do right now: study for the test, or play the game? If you're like many people, you might prefer the game. This is called **impulsivity**. Impulsivity means we want things that make us feel good and we want them now, and we want to

postpone things that take effort. Impulsivity causes problems, because it causes people to make unhealthy choices like overeating, procrastinating, smoking, overspending, and gambling. Now imagine something else. Imagine that it's *Wednesday* and your test is Friday. What should you do *tomorrow* (Thursday): study for the test, or play the game? Many people find it easy to decide to study tomorrow. But wait ... if we can make this good choice, maybe we're not so impulsive after all?

There's a mystery here. Do you want to study or play the game on Thursday? On Wednesday, you say you want to study on Thursday. On Thursday, you say you want to play video games on Thursday. It's almost like you're two different people who have different preferences and desires. The *Wednesday* you want to study on Thursday, but the *Thursday* you want to have fun. People often say they want to make good choices in the future, but instead they make impulsive choices when the future becomes now [1].

## WHAT IS PRECOMMITMENT AND HOW CAN IT HELP US MAKE GOOD CHOICES?

How can we break this cycle? At least since the time of the ancient Greeks, people have realized that the most powerful solution to impulsivity is something called **precommitment**. In the *Odyssey*, an adventure story written thousands of years ago, the main character, Odysseus, was trying to sail home from the Trojan war. To get home, he had to pass the island of the Sirens. The Sirens sang an alluring, magical song. Everyone who heard it sailed their ships toward the Sirens and straight into the island's rocks, sinking their ships. Odysseus wanted to hear the song, but he knew that if he brought his ship close enough to hear the song, he would also fall under its spell and drive his ship into the rocks. Odysseus thought of a way to beat the Sirens. He plugged his sailors' ears with wax, so they could not hear the song. He then ordered his men to tie him up, so he wouldn't take control of the boat and drive it into the rocks. This way, they sailed right past the Sirens. When Odysseus heard the song, he tried to break free from his ropes and steer the ship into the rocks, but the ropes held him. The ship safely sailed past.

What Odysseus did was an example of precommitment. Precommitment means to use control over your future self when you know your future self will make the wrong decision. Odysseus knew that his future self, hearing the Sirens' song, would want to sail into the rocks. So he took away the choice from his future self—literally by tying himself up!

One of the best ways scientists have found to study impulsivity is with an experiment called *the marshmallow test* [2]. In the marshmallow test, a kid (usually around 3–5 years old) is brought into a room with a single

### PRECOMMITMENT

Stopping your future self from making a bad decision. For example, locking up a chocolate cake so you can't get it later.

marshmallow on a plate. The kid sits down in front of the marshmallow and is told “If this marshmallow is still here in 20 minutes, I’ll give you two marshmallows. But if you eat this one marshmallow now, you won’t get the second marshmallow.” Then, the scientist leaves the room, leaving the kid alone with the marshmallow. Most children spend their time desperately trying to not eat the marshmallow, looking away, closing their eyes, smelling it but not tasting it, trying not to touch it or think about it. Just as with the impulsivity examples discussed at the start of this article, there are two selves in these children—one self wants to eat the marshmallow now, and the other wants to wait to get two marshmallows. Precommitment is like locking the marshmallow in a box before entering the room.

Precommitment is the best defense against impulsivity. The more we understand about precommitment, the more we can find effective new ways to help people make better decisions. However, most theories about how the brain makes decisions do not explain precommitment. To see why, we first need to know how the brain makes decisions.

## DECISIONS, VALUE, AND THE BRAIN

One of the most powerful theories about how people make decisions is that the brain has evolved to maximize the amount of **reward** it gets. What is reward? It’s things we like: tasty food, the approval of our friends, winning games, or a comfortable bed at night. Our brains evolved to help us get these things because in the course of evolution these things helped us survive and reproduce and make more brains (in our kids).

Through life, people get better and better at the skills that help them get the things they want. For example, this is why you go to school: to get the skills you need to get these rewards in life. But getting rewards can be very complicated. How does our brain learn how to get rewards? One of the brain’s clever tricks is to use something called reward **prediction errors** [3]. A reward prediction error is the difference between what you expected you would get and what you actually got. Here is an example. Imagine a candy machine where each candy bar costs \$1. If you put your \$1 in and get two candy bars out, you’re going to be more willing to put money in that machine—it gave you more candy than you expected. You had a prediction of how much candy you were going to get, and what you got was better than what you expected. This is a positive reward prediction error and it increases the likelihood of taking an action (putting more money in the machine). On the other hand, imagine that you put your \$1 in and get nothing out! Then, you’re going to be less willing to put money into that machine—it ate your dollar! This is a negative reward prediction error and it decreases the likelihood of taking an action. Of course, if you put your \$1 in and get your one candy bar out, you get what you expected and there is a zero or “flat” reward prediction error. Importantly, although you aren’t going to change how likely you are to take

### REWARD

The things we enjoy, like food, playing, being with friends, and the activities we like.

### PREDICTION ERROR

The difference between what you expect and what you get. It’s the same thing as “surprise.” Many parts of the brain learn by changing when they get prediction errors, so they can make better predictions next time.

that action, you still enjoy (hopefully!) the candy. Notice that pleasure and reward prediction error are different things.

This is how you can learn gradually from experience. If something happens as you expect, you don't need to learn anything. If you get more reward than you expected, you learn to do those actions more, and if you get less reward than you expected, you learn to do those actions less.

## PRECOMMITMENT IS A PUZZLE FOR SCIENTISTS

The theory of reward prediction errors is a very good one. A lot of data that have been measured from the brain can be explained by this theory. However, it has a serious problem: it can't explain precommitment.

That's because precommitment is only *necessary* if you prefer video games when I ask you on Thursday, and precommitment is only *possible* if you prefer studying when I ask you on Wednesday. But learning from reward prediction errors doesn't produce different preferences at different times. That theory would say that if you prefer video games on Thursday, you would prefer video games on Wednesday as well. Learning from reward prediction errors teaches you the value of each thing, and those values stay the same relative to each other. If you learn that something is better in the future, then you've also learned that it's better in the present.

## A POSSIBLE SOLUTION

We recently suggested a solution to the puzzle of how precommitment happens in the brain [4]. Different parts of the brain might actually learn different answers about how good actions are. These different parts of the brain can therefore have different preferences about which things they like best. And this is very important: we also suggest that these parts of the brain think about the future very differently. Some parts immediately lose interest in a reward if they can't get it immediately, and others are willing to wait a long time. This has a surprising effect: it allows you to prefer different things at different times.

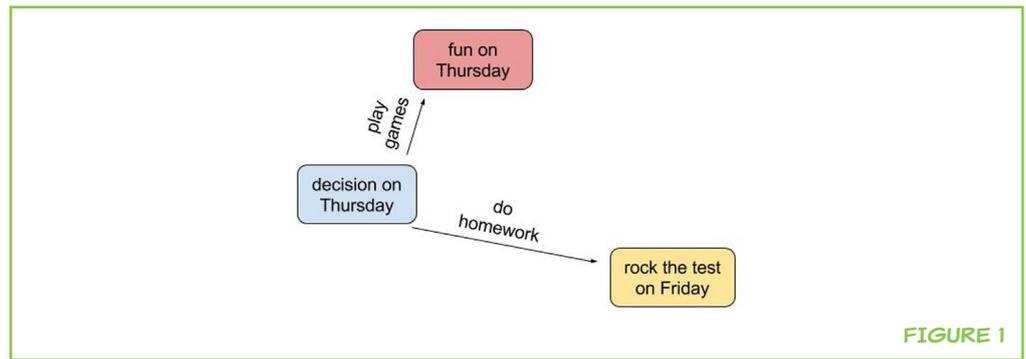
To understand how this works, let's take a careful look at what precommitment really is. Scientists studying decision-making describe decisions as mentally moving through a set of "states of the world" or "situations." These situations describe how the world looks at a specific moment in time. And your decisions determine how you move between these situations—like a flow chart. For example, Figure 1 shows a choice between a small reward you can get now versus a large reward that you have to wait for.

What does this have to do with precommitment? Let's see if we can define precommitment using this same kind of flow chart. Take a look at Figure 2.

**FIGURE 1**

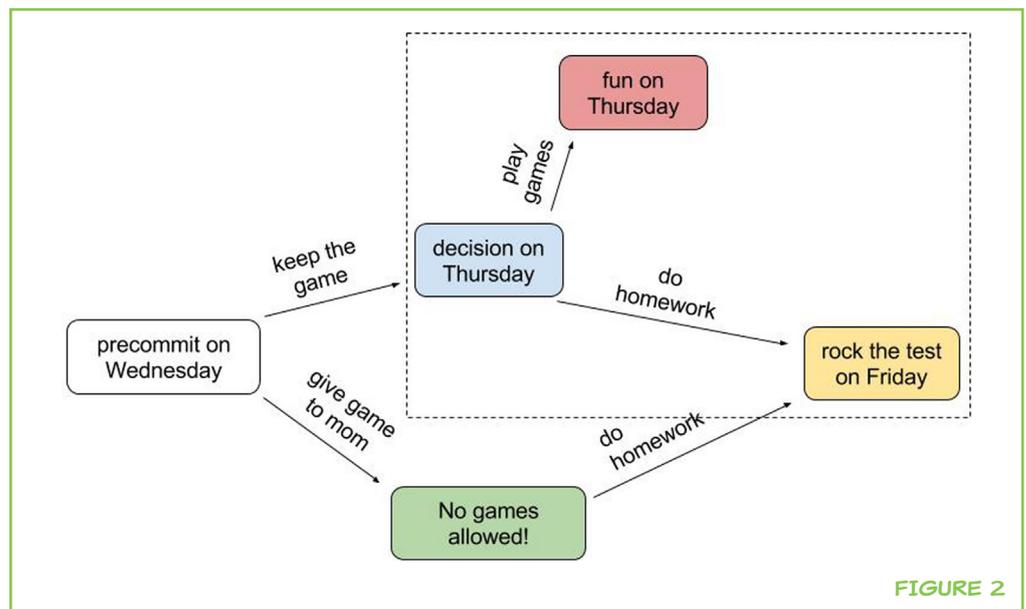
Flow chart of a single decision.

On Thursday, you have two choices: one provides a small reward (the fun of playing games) immediately, while the other provides a bigger reward (doing well in school), but you have to wait until Friday to get that reward.

**FIGURE 1****FIGURE 2**

A flow chart showing the possibility of precommitment.

Inside the dashed box is exactly the same choice from Figure 1. But now we've added a "pre-choice" on Wednesday—do you want to give the video game to your mom for safekeeping? If you *don't* give it to her, then on Thursday you'll be faced with exactly the choice in Figure 1. But if you *do* give it to her, you're guaranteed not to be tempted. Giving the game to your mom is what we call precommitment.

**FIGURE 2**

In our video game versus studying example, you know that on Thursday you're going to be tempted to play games instead of studying. So on Wednesday, it might be a good idea to give the game to your mom, and tell her not to let you have it back until after the test. The interesting thing is that, during the delay between Wednesday and Thursday, your preference changes from studying to video games.

Describing the problem in this way gives us a new way to look at the problem. It tells us that our decision systems prefer video games on Thursday, but prefer studying on Wednesday. Fortunately, our Wednesday self can precommit, to protect us from our Thursday self's preferences.

**TWO (OR MORE!) SELVES**

In our theory, we proposed that people actually have multiple opinions about the reward at each point in time. This is called a *distributed* representation

of reward. In our theory, the brain contains lots of different opinions—it's like a group of friends deciding what to do.

Imagine two friends, Bart and Lisa (Figure 3). Lisa is very patient, and Bart is very impulsive. Importantly, we will assume that both Bart and Lisa learn by reward prediction errors, so each of them always prefers either studying or video games. Let's say Bart always prefers games over studying, by a lot. Lisa, on the other hand, always prefers studying over games.

Bart is impulsive, which means he cares a lot about the *present*. So on Thursday, his desire to play video games is very large. But because he's so impulsive, on Wednesday this desire is very small. Meanwhile, Lisa's values don't change much over time. For her, the future is nearly as important as the present. So on Thursday she prefers studying to video games, and on Wednesday she *still* prefers studying to video games.

If they are trying to make a decision together about what to do, then on Thursday, Bart's strong preference for video games will dominate the conversation and if we were to take the average of Bart and Lisa's preferences, video games would win out over studying, but on Wednesday, Lisa's low impulsivity will dominate the conversation and if we took the average, studying would win out. If you average Bart and Lisa's preferences together, the average completely reverses what it wants to do between Wednesday and Thursday. Even though neither of them on their own could reverse what they prefer and precommit, together they can. Now, think of Bart and Lisa as two voices inside your own brain.

## WHAT DOES ALL OF THIS MEAN?

This theory of multiple selves means that it might be possible to activate the right selves at the right time, to help us make better decisions. There is evidence from brain scans that tells us that changes in diet might help tip this balance [5]. It might also be possible to practice learning to listen to our

### FIGURE 3

Finding your inner Bart and your inner Lisa. Lisa always prefers studying, as indicated by the pictures in the Wednesday and Thursday columns, where the pictures of the books are larger than the pictures of the video game. Bart always prefers video games, indicated by the larger picture of the video game on Wednesday and Thursday. But their *average* preference shifts from studying to video games as time passes, as shown at the bottom of the figure.

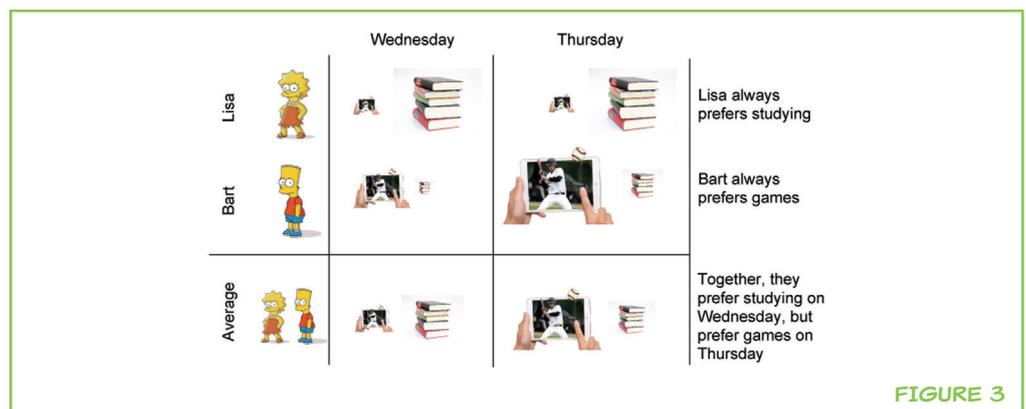


FIGURE 3

more patient selves. Or we might find tricks or strategies that fit with our unique mixture of patience versus impulsivity, like the earlier example of giving mom the video game on Wednesday so you can't play on Thursday. As you understand yourself better, you can learn what works to help you avoid impulsive decisions.

If you know that you are going to make an impulsive decision, you can use the multiple selves that you are to prevent yourself from falling into that temptation. You can lock the marshmallow away or tie yourself to the mast like Odysseus. All you need to know is that it's OK to realize you will be tempted to make an impulsive decision and to take the action earlier to prevent it.

## ORIGINAL ARTICLE REFERENCE

Kurth-Nelson, Z., and Redish, A. D. 2012. Don't let me do that! – models of precommitment. *Front. Neurosci.* 6:138. doi:10.3389/fnins.2012.00138

## REFERENCES

1. Ainslie, G. 2001. *Breakdown of Will*. Cambridge, UK: Cambridge University Press.
2. Mischel, W., Ebbesen, E. B., and Raskoff Zeiss, A. 1972. Cognitive and attentional mechanisms in delay of gratification. *J. Pers. Soc. Psychol.* 21(2):204. doi:10.1037/h0032198
3. Montague, P. R., Dayan, P., and Sejnowski, T. J. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16(5):1936–47.
4. Kurth-Nelson, Z., and Redish, A. D. 2012. Don't let me do that! – models of precommitment. *Front. Neurosci.* 6:138. doi:10.3389/fnins.2012.00138
5. Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., and Yamawaki, S. 2004. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.* 7(8):887–93. doi:10.1038/nn1279

**SUBMITTED:** 05 January 2017; **ACCEPTED:** 31 May 2017;

**PUBLISHED ONLINE:** 16 June 2017.

**EDITED BY:** Beatriz Luna, University of Pittsburgh, United States

**CITATION:** Kurth-Nelson and Redish (2017) Precommitment: A Way around Temptation. *Front. Young Minds* 5:26. doi:10.3389/frym.2017.00026

**CONFLICT OF INTEREST STATEMENT:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**COPYRIGHT** © 2017 Kurth-Nelson and Redish. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original

author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## REVIEWED BY



### **GABI, 11 YEARS OLD**

I am in sixth grade and have three siblings. My favorite subjects in school are English and Math. I love to read mystery novels, traveling, and tumbling. I also like to swim and tumble on the trampoline.

## AUTHORS



### **ZEB KURTH-NELSON**

Zeb Kurth-Nelson is a research scientist at DeepMind. Previously he did postdocs at the University of Minnesota and University College London, where he studied the interplay between representation and decision processes and how they function differently in psychiatric disorders. His doctoral work at the University of Minnesota focused on glia-to-glia and glia-to-vascular purinergic signaling in the mammalian retina.



### **A. DAVID REDISH**

David Redish is a professor in the Department of Neuroscience at the University of Minnesota. He is a poet, a playwright, and a neuroscientist. He is particularly interested in the human condition and in how humans (and other animals) make decisions. He is the author of several books, including *The Mind within the Brain: How we make decisions and how those decisions go wrong*. \*redish@umn.edu