

# Cooperate Without Looking

Suppose a friend asks you to proofread a paper...

You **hesitate** while thinking about how big a pain it is and say, “Hmm. Um. Well, OK.”

You’re a jerk even though you agreed.

Colleague asks you to attend his talk.

You **ask**, “how long will it last?” before agreeing to attend.

You’re a jerk, whether you end up attending or not.

In the first example, you hesitated as you deliberated over whether it was worthwhile

In the second example, you asked for additional information as you deliberated

Both reduce “kudos” for cooperation

Why?

# Can reciprocity explain this?



Image courtesy of [kjelljoran](#) on Flickr. CC BY-NC-SA

In reciprocity, our behavior depends on others' actions, not thought/search processes

Therefore, reciprocity cannot explain reduced kudos from deliberating

So what is it?

Our intuition...

Cooperators who don't deliberate, or "look", can be trusted to cooperate even when the temptation to defect is high

Intuition is not enough...

Is the extra trust gained from not looking worth the occasional high cost?

I.e., can cooperating without looking (CWOL) be sustained as a Nash Equilibrium?

Can we get some new predictions or prescriptions?

When do we expect people to care if their colleagues “look”?

When do we expect colleagues to avoid looking?

Given that people don't consciously choose whether to look, can this behavior arise from learning dynamics?

To deal with these question, we need formal analysis

Proceed as follows...

- 1) Formally model this using “the envelope game”
- 2) Show “not looking” is an equilibrium in this game
- 3) Use this to find (simple, intuitive) conditions under which to expect “care if look” and “avoid looking”
- 4) Show “not looking” emerges in replicator dynamic

Then use envelope game to gain insights into less straightforward social phenomena:

- Flip-flopping politicians and “principled” people
- Why religions reward anonymous giving
- Taboo trade-offs
- Love

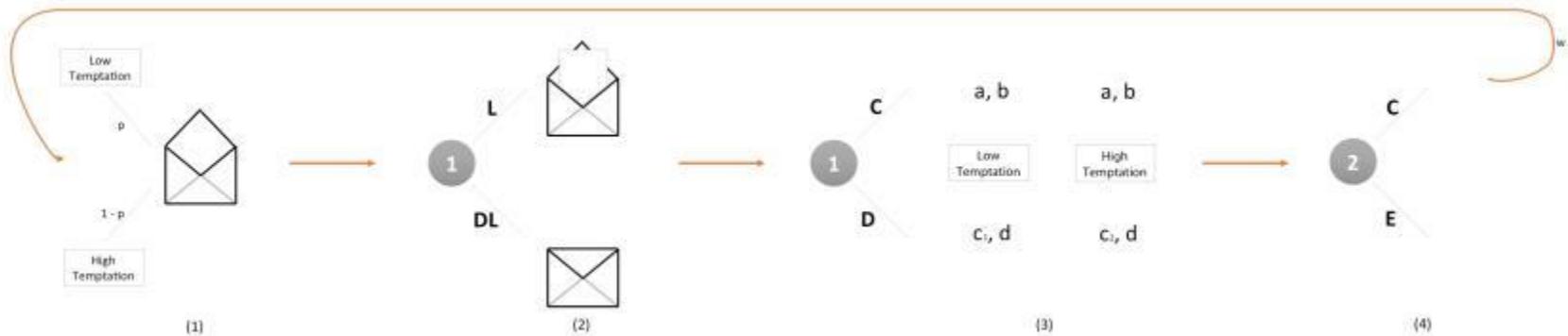
Let's start with the model...

# “The Envelope Game”

Two players

Play a repeated game

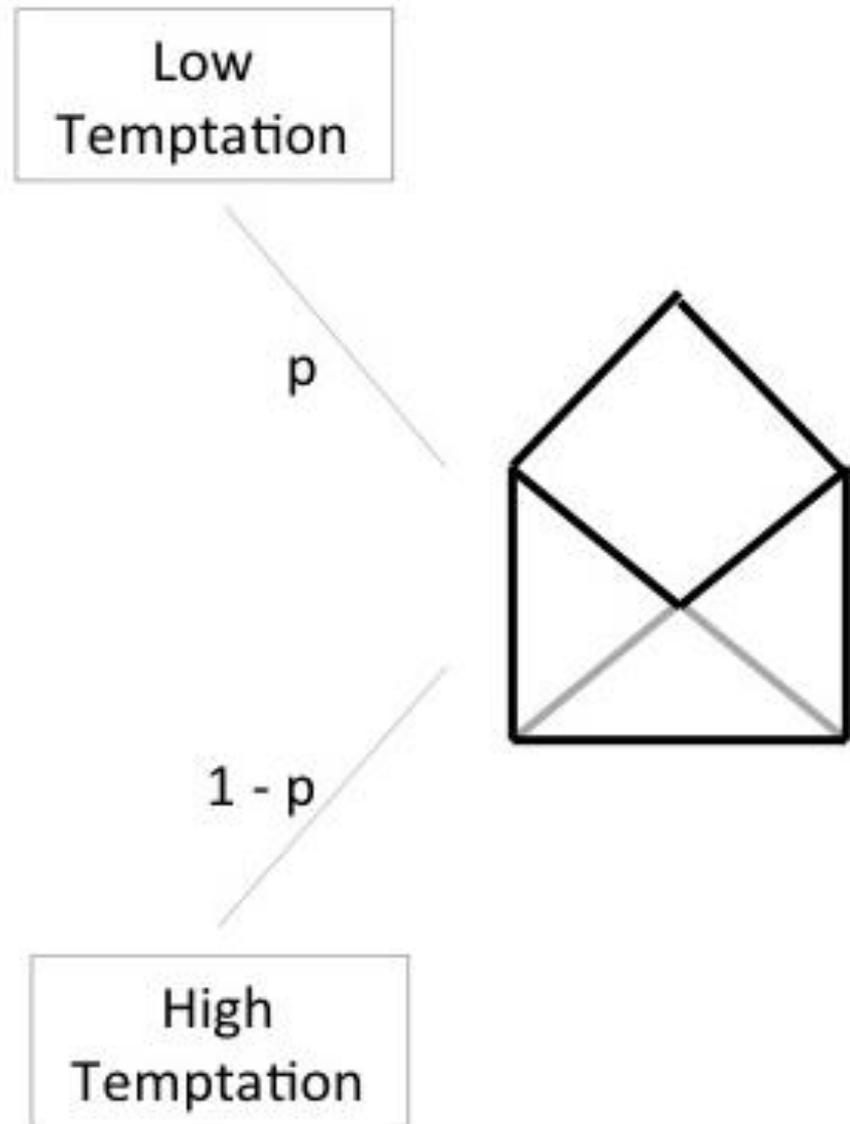
Each “stage” has four parts

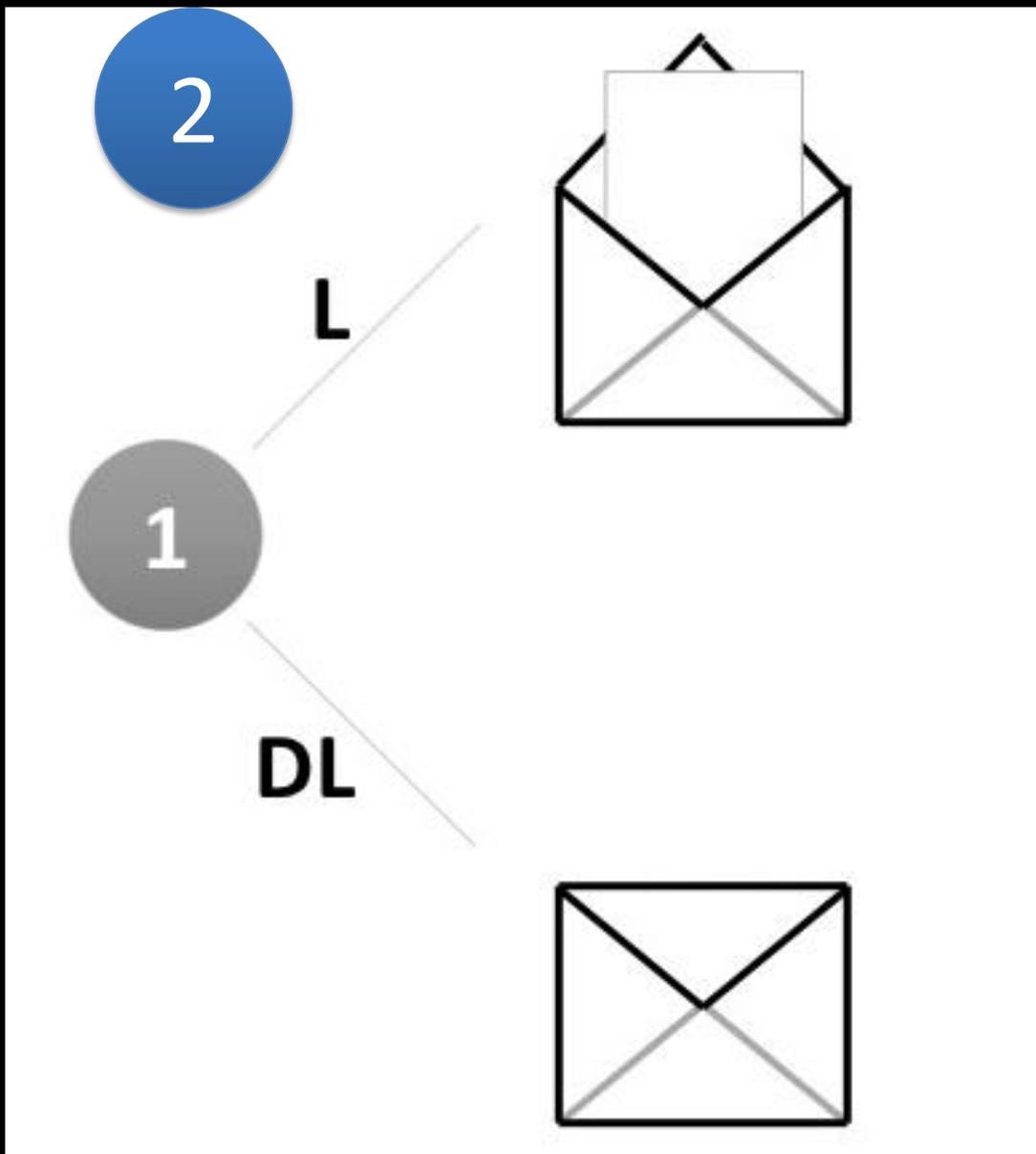


First...

We model variation  
in costs of cooperation

With probability  $p$  Low  
Temptation “card”  
is chosen and stuffed in  
envelope





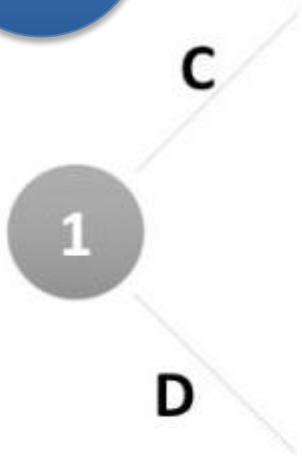
Second...

We model player 1's choice of whether to "look"

1 chooses whether or not to open the envelope

Crucially we assume others (player 2) can observe whether 1 looks

2

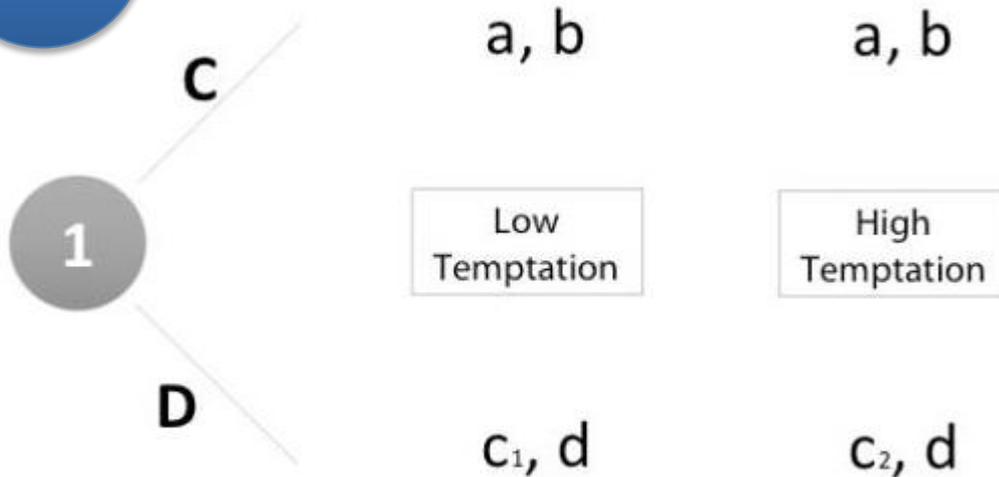


Third...

1 then chooses  
whether or not to  
Cooperate

2 is again able to  
observe

2



And players receive payoffs which depend on the temptation level and the choice made by 1

Assumptions about payoffs:

Both prefer Cooperative interaction to no interaction ( $a > 0, b > 0$ )

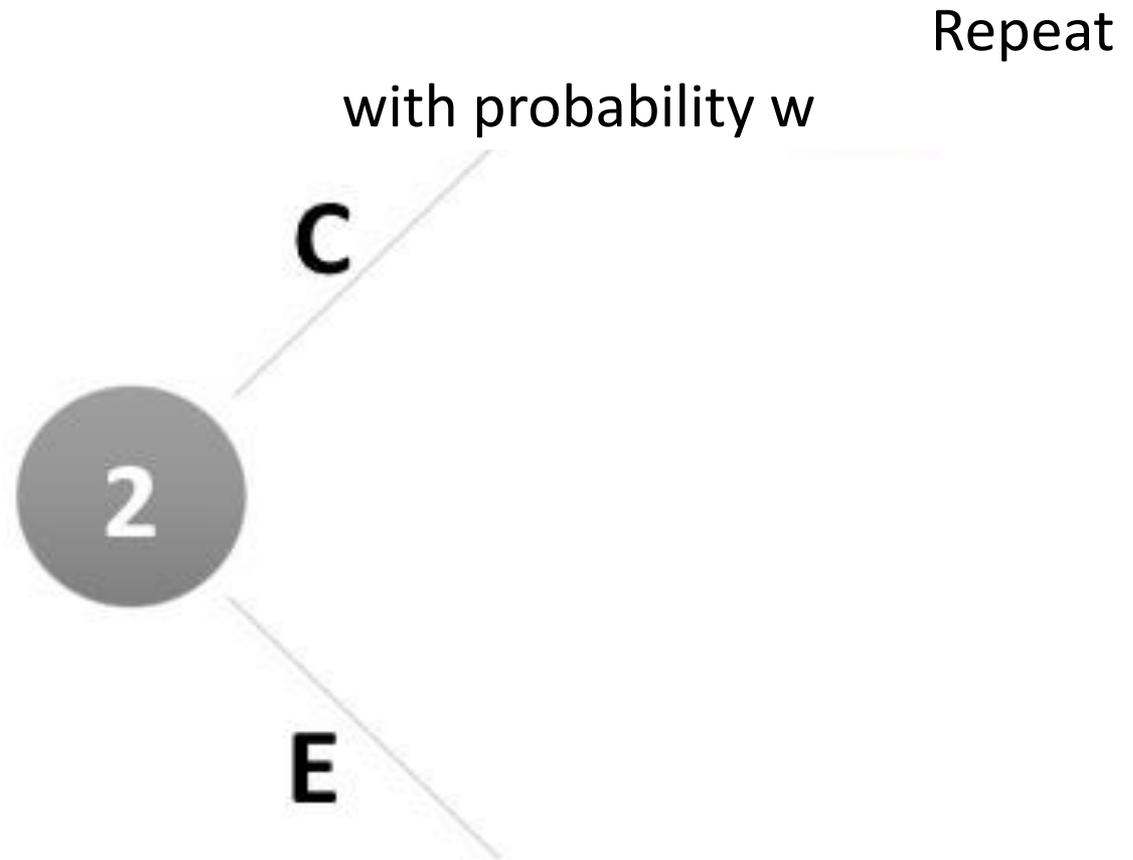
Cooperation is costly to 1 ( $c_1 > a$ ) and beneficial to 2 ( $b > d$ )

1 gains more from defecting in "high temptation" ( $c_2 > c_1$ )

Fourth and finally...

We model others' "trust" in player 1

Player 2 chooses whether or not to repeat the interaction, with discount  $w$



Let's take a second to think about strategies in this game...

A strategy for player 1 identifies whether she looks and whether she cooperates in each round as a function of what has happened in the past

E.g., In even rounds, don't look, then cooperate. In odd rounds look and defect

A strategy for player 2 identifies whether she continues or exits in each round as a function of what has happened in the past

E.g., Continue if player 1 has never looked

And payoffs?

The ones we showed you are for one round. We need to “add them up” across rounds

E.g., If player 1 look and cooperates when temptation is low and player 2 continues if player 1 cooperates

Player 1's payoffs:  $a/(1 - w)$

Player 2's payoffs:  $b/(1 - w)$

Another example...

Suppose player 1 always defects and player 2 always exits

Player 1's payoffs:  $c_1p + c_2(1-p)$

Player 2's payoffs:  $d$

Let's also take a second to think about the key assumptions...

Can avoid thinking about/gathering info about costs  
Others can detect

How can others detect?

1's reaction time?  
The questions 1 asks 2?

One more important assumption:

Defection is so costly to player 2 that it's not worth interacting even if he only expects player 1 to only defect when the temptation is high

$$bp + d(1-p) < 0$$

Now ready to discuss results...

Main result:

1 Cooperates without looking (CWOL)

2 continues if 1 CWOL

is an equilibrium

when  $a/(1-w) > c_1p + c_2(1-p)$

Sketch of proof:

For player 1, will not deviate (look or defect) as long as payoff from future interaction is greater than *expected* temptation,  $a/(1-w) > c_1p + c_2(1-p)$

For player 2, if 1 isn't looking, is no better off if stops attending to looking and worse off if ends interaction

So in response to our first set of questions...

Is the extra trust gained from not looking worth the occasional high cost?

I.e., can cooperating without looking (CWOL) be sustained as a Nash Equilibrium?

The answer is yes

Next we'd like to interpret the equilibrium condition to answer our next set of questions...

Can we get some new predictions or prescriptions?

When do we expect people to care if their colleagues "look"?

When do we expect colleagues to avoid looking?

But first...

There is another equilibrium where...

1 looks

2 cooperates if 1 cooperates (CWL)

when  $a/(1-w) > c_2$

Let's compare these...

CWOL is an equilibrium when

$$a/(1-w) > c_1p + c_2(1-p)$$

CWL is an equilibrium when

$$a/(1-w) > c_2$$

Since  $c_2 > c_1$ , CWOL is an equilibrium for a strictly larger range of  $a$ 's

Combine the conditions above to get the condition where CWOL is an equilibrium but CWL isn't:

$$c_1 p + c_2(1 - p) < a/(1-w) < c_2$$

This answers our questions: 2 will care if 1 looks and 1 will avoid looking when the **expected gains from defecting** are small but the **maximal gains** are large

It also teaches us that the ability to avoid looking and to detect looking enables cooperation when otherwise wouldn't be possible

Summary so far...

We have shown CWOL is an equilibrium

And interpreted equilibrium condition to learn when not looking and attending to looking are important

But no one is consciously choosing to not look or to attend to looking

Instead, use feelings, heuristics or ideologies which guide them when to look and when to attend to looking

I.e.,

I respect “principled” people

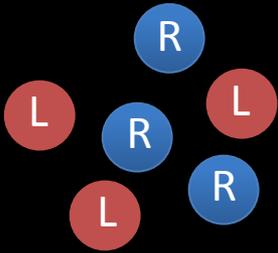
I can’t fathom trading lives for money

Without rationality, need another justification  
for Nash

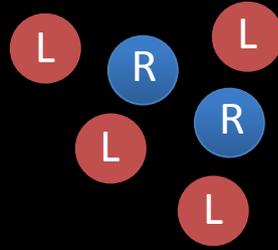
So we use dynamics

As usual, only assumption... more successful strategies become more frequent

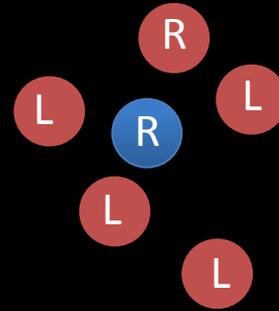
(e.g. because more likely to be imitated)



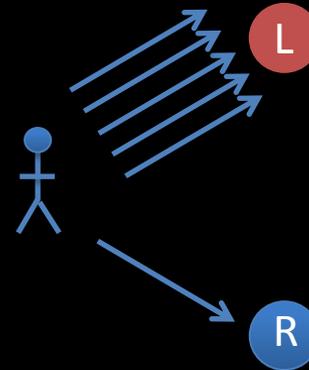
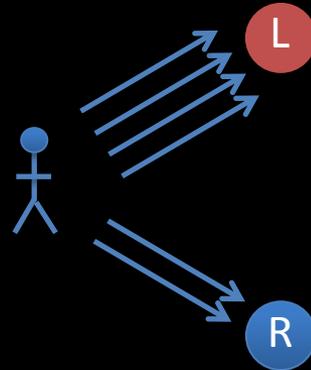
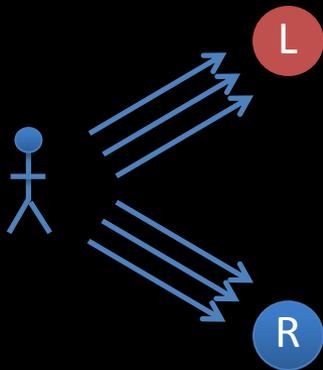
T=0



T=1



T=2



Of course *strategies* aren't imitated

Feelings/heuristics/ethics/ideologies are

But these are just ways of implementing strategies, so we represent them by the strategy they implement

The dynamic we use is replicator dynamic

For replicator, we need to simplify strategy space...

For each player, we use the following strategies

Player 1	Player 2		
	Continue if Player 1 Cooperates Without Looking	Continue if Player 1 Cooperates	Always Exit
Cooperate Without Looking	$\frac{a}{1-w}, \frac{b}{1-w}$	$\frac{a}{1-w}, \frac{b}{1-w}$	$a, b$
Cooperate With Looking	$a, b$	$\frac{a}{1-w}, \frac{b}{1-w}$	$a, b$
Look and Cooperate Only if Temptation is Low	$ap + c_2(1-p), bp + d(1-p)$	$\frac{ap+c_2(1-p)}{1-pw}, \frac{bp+d(1-p)}{1-pw}$	$ap + c_2(1-p), bp + d(1-p)$
Always Defect	$c_1p + c_2(1-p), d$	$c_1p + c_2(1-p), d$	$c_1p + c_2(1-p), d$

(b)

We also need to be a little more complete in classifying equilibria

Here are the two we identified before

There's also a third one which is always an equilibrium

Player 1	Player 2		
	Continue if Player 1 Cooperates Without Looking	Continue if Player 1 Cooperates	Always Exit
Cooperate Without Looking	$\frac{a}{1-w}, \frac{b}{1-w}$	$\frac{a}{1-w}, \frac{b}{1-w}$	$a, b$
Cooperate With Looking	$a, b$	$\frac{a}{1-w}, \frac{b}{1-w}$	$a, b$
Look and Cooperate Only if Temptation is Low	$ap + c_2(1-p), bp + d(1-p)$	$\frac{ap+c_2(1-p)}{1-pw}, \frac{bp+d(1-p)}{1-pw}$	$ap + c_2(1-p), bp + d(1-p)$
Always Defect	$c_1p + c_2(1-p), d$	$c_1p + c_2(1-p), d$	$c_1p + c_2(1-p), d$

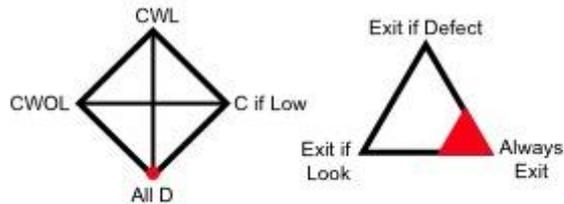
(b)

## Equilibrium Classification

Player 1

Player 2

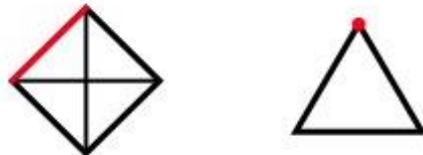
All D:



CWOL:



CWL:



Now we lump together ones which are “behaviorally identical”

For example, in CWOL, as long as enough 2s exit if *look* it doesn't matter if some 2s exit if *defect*

Here's how the simulation works

For each of many parameter values...

For each of 5,000 trials...

We seed the population with random mixtures of strategies

Then wait for the population to stabilize

Then classify the outcome

We find...

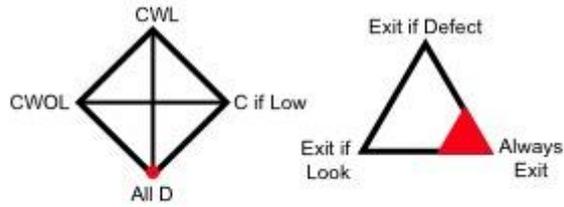
Population ends up at CWOL fairly often in parameter region where it is an equilibrium

# Equilibrium Classification

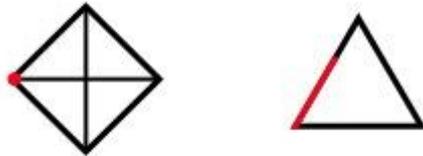
Player 1

Player 2

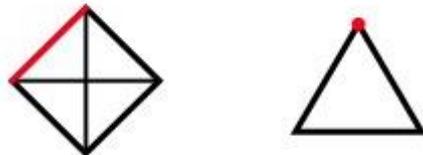
All D:



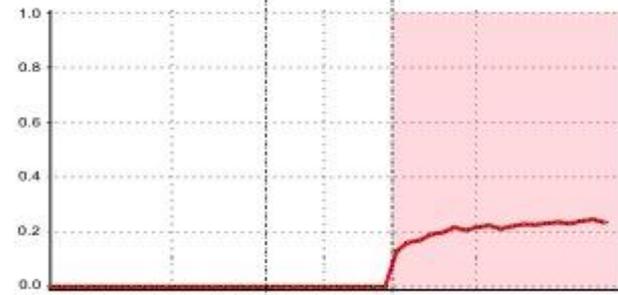
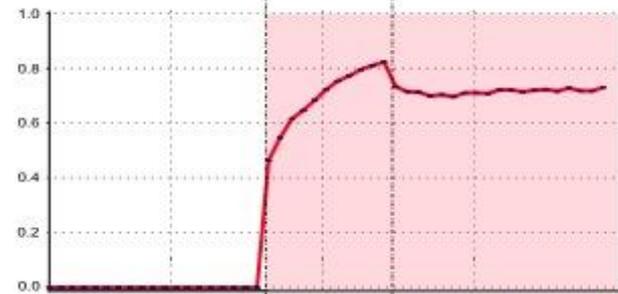
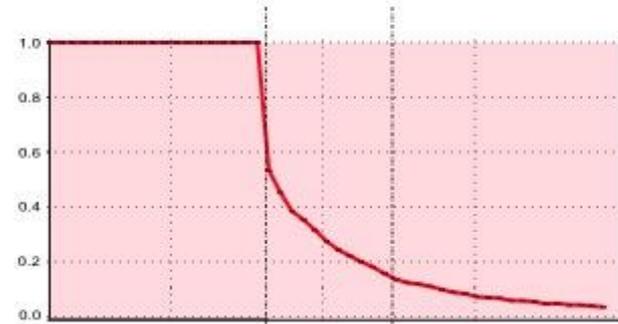
CWOL:



CWL:



# Frequency



a

And that answered our third set of questions

Given that people don't consciously choose whether to look, can this behavior arise in dynamics?

The answer is yes

Let's apply these results to a few interesting social phenomena...

First one... “flip-flopping” politicians



Image is in the public domain.

**I was for  
the military  
before I was  
against it.**

Why is “flip-flopping” bad?

Isn't it better for Kerry to be sensitive to constituents' opinion?

This is part of a general question...

Kerry decides policy “strategically” and not based on “principles”

Why do we like people who have “principles” and not those who are “strategic”?

And when will we particularly care?

CWOL provides the following possible explanation...

Kerry (re)decides political stance after calculating costs to his career

So he cannot be trusted to make good policy when it would hurt his career

This generalizes to strategic vs. principled people more generally

And we can use our equilibrium condition to identify when we will especially care that people are principled

When usually have fairly aligned incentives but are rare occasions that could benefit them and harm us

E.g., crucial that girlfriend/boyfriend is principled, but not that drinking buddy is

Any evidence?

Second puzzle... Why do religions reward anonymous giving?

For example...

# Maimonides' Eight Levels of Giving

8. He who does not give enough—and even that unwillingly.

7. He who does not give enough—but what he gives, he gives graciously.

6. He who gives enough—but only after he is asked.

5. He who gives enough and before he is asked—but with both parties knowing each other.

4. He who gives and does not know who receives—but the recipient knows who gave it.

3. He who gives and knows who will receive it—but the recipients will not know who gave it.

2. He who gives without knowing who will receive—and without the recipient knowing who gave it.

1. He who gives—and also assists the recipient in addressing the reason for the need in the first place.

Do you know of examples from other religions?

Why reward anonymity and not just giving?

An anonymous gift can feed no more starving children

More generally,

Why do we admire givers more when they give anonymously?

For example...

See article about [Steve Jobs' philanthropy](#).

Source: UK Daily Mail, May 24, 2013

We argue that anonymous giving is more honorable because such people are giving without “looking” at the reputational benefits

So they can be expected to continue giving, even when the reputational effects are minimal

(Note here it is benefits not temptations, but works the same)

But people typically know they're giving  
anonymously

Might be guided by principles to do so anyway

Suggests trade-off between giving so everyone knows and admires a little, and a few people might find out and admire a lot

Third puzzle... taboo tradeoffs

What is a taboo trade-off?

Let's illustrate with an example...

A third of Medicare dollars are spent in the last month of life

(Source: <http://www.forbes.com/sites/michaelbell/2013/01/10/why-5-of-patients-create-50-of-health-care-costs/>)

Presumably, for a lot of folks, we know those dollars are unlikely to help

And those dollars could be used to more effect elsewhere... for example, to improve the other 959 months of our lives

But if you ask a doctor, they can't even imagine making this trade-off

As Tetlock and Roth point out...

We find it “disgusting” to even consider trading  
off money against life

I.e., it’s “taboo”

Note that this taboo is governed by emotions or ideologies—most people have not consciously considered it

And that it is upheld despite enormous costs

Suggests psychologically deep

This generalizes...

We find it disgusting to even consider trading off money with many “sacred” goods

We ask:

Why?

Which goods are liable to be sacred?

Should we, as a society, respect taboos?

Here's our suggestion for why and when...

We find such tradeoffs disgusting because they signal a willingness to look at the benefits of defecting...

... in domains where rare instances of huge potential temptations

E.g.,

- If partner considers selling sex, then might cheat when the right person comes along
- If caretaker put a price tag on life, might compromise it for especially lucrative offer

Policy implication:

While politicians might want to signal that they would never trade lives for money

Gains from appearing trustworthy accrue to the politicians while costs accrue to us

We need a government that surmounts taboos to tackle these admittedly hard-to-fathom questions

Any evidence to support this claim?

How about evidence to support our prediction on which items will be sacred?

Last puzzle... Intuitive cooperation

First piece of background...

Psychologists sometimes discuss two “types” of decisions:

System 1: intuitive

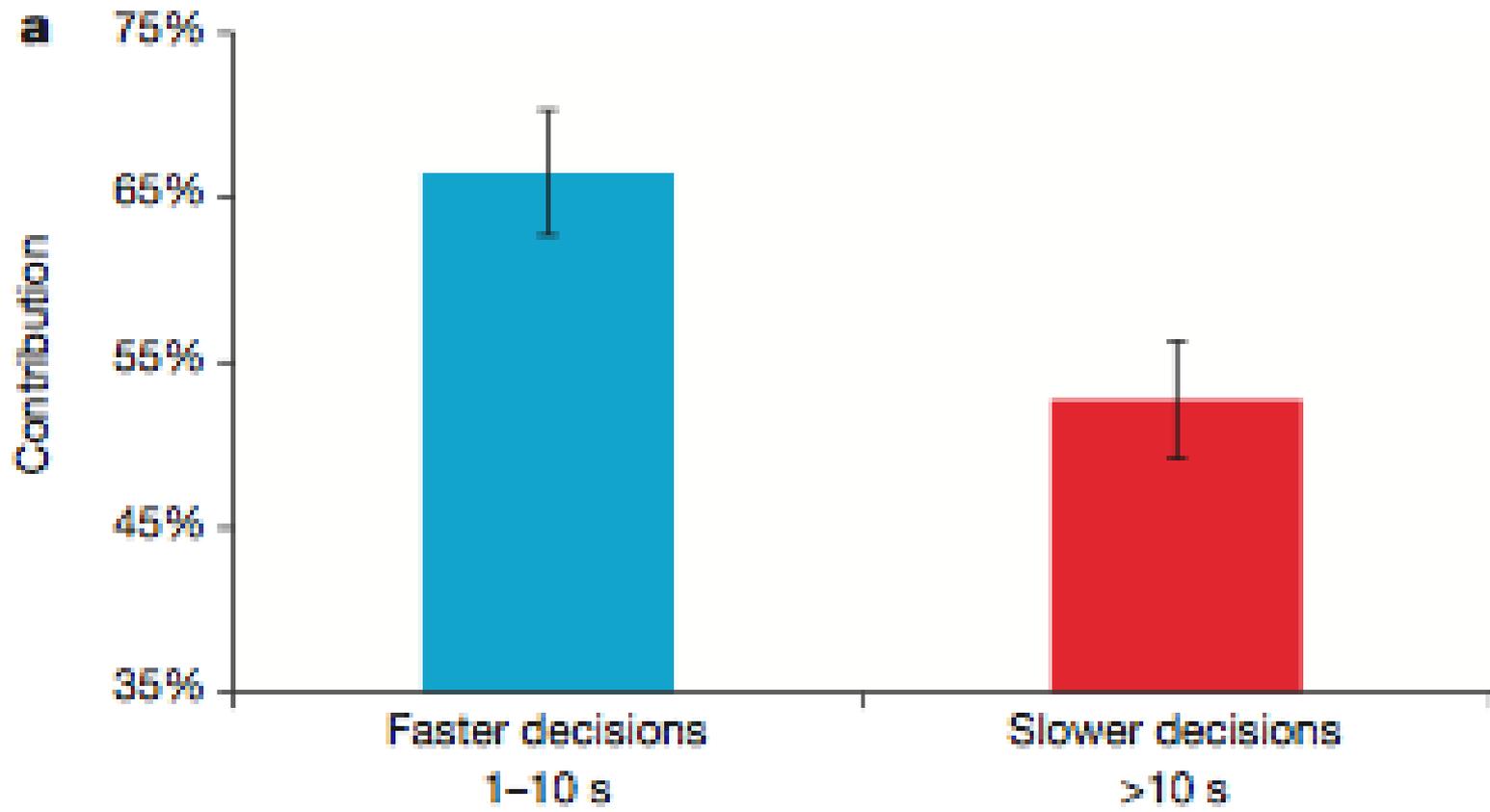
System 2: deliberate

Second piece of background...

People cooperate intuitively

Cooperators have faster response time  
than defectors

When forced to slow down, cooperate less



Reprinted by permission from Macmillan Publishers Ltd: Nature.  
Source: Rand, David G., Joshua D. Greene, et al. "Spontaneous Giving and Calculated Greed." *Nature* 489 (2012): 427-30. © 2012.

Source: Rand et al. 2012

Why would people cooperate intuitively?

Our model suggests that intuitive cooperation might be a way of not looking

If some lab subjects not looking, and some looking, then, on average, cooperators decide faster

Gives two novel predictions:

Especially likely to be intuitive in cooperative situations

Others will prefer to interact with intuitive than deliberative cooperators

Any ideas for how to test?

Let's summarize...

People attend to whether cooperators deliberate.  
Reciprocity can't explain this

We use simple model and dynamics to show that when  
can detect looking, sometimes worth not looking

That sometimes: when relationship is beneficial on  
average but there are occasional large temptations

Possibly explains puzzling social behaviors

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.11 Insights from Game Theory into Social Behavior  
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.