Cooperating Without Looking

Moshe Hoffman,[1,2*†] Erez Yoeli,[2,3*] Martin Nowak[2]

[1]Department of Computer Science and Engineering,
University of California at San Diego, La Jolla, CA 92093
[2]Program for Evolutionary Dynamics,
Harvard University, Cambridge, MA 02138
[2]Federal Trade Commission,
600 Pennsylvania Ave. NW, Washington, DC 20004
*These authors contributed equally to this work.
†To whom correspondence should be addressed; E-mail: ███████████████████.

Cooperation occurs when we take on costs to help others. A key
mechanism by which cooperation is sustained is reciprocity: indi-
viduals cooperate with those who have cooperated in the past. In
reality, we not only condition on others' past cooperative actions,
but also on the decision making process that leads to cooperation:
we trust more those who cooperate without calculating the costs
because they will cooperate even when those costs are high. We
propose a game theory model to explain this phenomenon. In our
model, player 1 chooses whether or not to cooperate with player 2.
Player 1 faces a stochastic temptation to defect and, before choos-
ing whether to cooperate, also decides whether to "look" at the
realized temptation. Player 2 observes not only whether player 1
ultimately cooperated but also whether she looked, then decides

whether or not to continue interacting with player 1. We find conditions in which there is an equilibrium where player 2 chooses to interact with player 1 only if player 1 cooperates without looking (CWOL) and player 1 chooses to CWOL. We show that this equilibrium is robust to both high degrees of rationality, as modeled by subgame perfection and learning or evolutionary dynamics, as modeled by the replicator dynamic. Using computer simulations, we also show that it emerges with high frequency, even in the presence of other equilibria. Additionally, we show that the ability for player 1 to avoid looking, and the ability for player 2 to detect looking increases cooperation. We propose this model as a possible explanation for a number of interesting phenomena, and thereby derive novel predictions about these phenomena, including why we dislike "flip-flopping" politicians and respect principled people more generally, why people cooperate intuitively, and why people feel disgust when considering taboo trade-offs, and why people fall in love.

Cooperation occurs when we take on costs to help others. A key mechanism by which cooperation is sustained is reciprocity: individuals condition their behavior on others' past actions [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. In reality, we not only condition on others' past actions, but also on the decision-making process that lead to cooperation: we place more trust in cooperators who do not take time to carefully weigh the costs of cooperation and who do not try to collect data on such costs before deciding whether to cooperate. For example, we are impressed by colleagues who agree to proofread a paper without thinking twice, and view with suspicion those who ask, "how long will it take?" before agreeing to attend a practice talk. Such considerations are left out of standard

2

models of reciprocity, which only attend to cooperative actions and not the deliberation process leading up to the action.

We develop a simple model to explain why "looking" at the costs of cooperation is viewed with suspicion. The explanation we suggest is quite intuitive: those who cooperate without looking (CWOL) can be trusted to cooperate even in times when there are temptations to defect. While this insight can be captured without the need for a formal model, it is less clear that cooperators will choose to not look, since they pay a price by cooperating blindly in tempting circumstances. Moreover, the formal model, as will be seen, helps explicate when CWOL will occur as well as what difference the ability to not look and to observe others not looking will make.

We formalize this idea using what we call the envelope game (see figure 1). The envelope game distills an interaction between two individuals, or players, in an uncertain environment. Thus, we start by assuming there is a distribution of payoffs with two possibilities: one in which defection is relatively tempting, and another in which it is not. The temptation to defect is randomly determined. Defection is not tempting with probability p and tempting with probability $1 - p$. Both players know how likely it is that defection is tempting. However, at this point, neither player knows size of the temptation. That is, the temptation is placed inside an envelope and the envelope is sealed without the players knowing its content. Next, we assume that one of the players, player 1, chooses whether to learn the size of the temptation, either via mental deliberation or by gathering information. We model this in a simplified way by assuming that player 1 has a dichotomous choice: she can choose to open the envelope and look inside it, or not. If she opens the envelope and looks, she learns the size of the temptation. If she does not open the envelope and look, she only knows the distribution of payoffs. Player 1 then chooses whether to cooperate or defect. We model cooperation as costly to player 1, regardless

3

of the temptation, but more costly if the temptation is high. Thus, if the temptation is low, player 1 gets $a > 0$ if she cooperates and $c_l > a$ if she defects. If the temptation is high, she gets a if she cooperates and $c_h > c_l$ if she defects. We assume that cooperation benefits the other player, which we refer to as player 2. We assume, for simplicity, that player 2 gets $b > 0$ if 1 cooperates and $d < 0$ if 1 defects, and that these payoffs to player 2 are the same regardless of 1's temptation. For now, we also assume that $d < p/(1-p) \cdot b$; that is, defection is sufficiently harmful to player 2 that player 2 prefers to avoid the interaction if player 1 only cooperates some of the time. Finally, we model player 2's trust in player 1. We model this by giving player 2 the choice of whether to continue the interaction with player 1. If player 2 continues the interaction, then with probability w, all previous steps are repeated, potentially indefinitely. That is, the temptation to defect is randomly chosen again, player 1 chooses whether to look at it, and so on. With probability $1 - w$, the game ends. The parameter w can be interpreted as the likelihood of a future interaction or the player's discounting of future payoffs. Crucially, we assume that player 2 observes both of player 1's choices: not only whether 1 cooperated, but also whether 1 first looked. Our model, therefore, applies to situations where player 1's looking is somewhat observable, for example, if player 1 can gain additional information about the costs of defection by asking player 2 questions, or if player 1 can take time to ponder the decision and player 2 can observe player 1's reaction time. We assume that players maximize their expected payoffs. Note that since we assumed $a > 0$ and $b > 0$, both players benefit from a cooperative interaction.

In the envelope game, a strategy for player 1 dictates whether she will look and whether she will cooperate, as a function of whether she has looked and cooperated in the past, and as a function of the temptation if she has just looked. For example, one relatively simple strategy for player 1 is to look and cooperate whenever the temptation is low.

Similarly, a strategy for player 2 dictates whether he will continue or exit as a function of everything player 1 has done in the past. One relatively simple strategy for player 2 is to continue so long as player 1 has cooperated. If both players play according to this pair of strategies, each time the temptation is low, player 1 cooperates and gets a, player 2 gets b, and the game continues with probability w until the first time the temptation is high. Then, player 1 defects and gets c_h and player 2 gets d. Some straightforward calculations yield the players' expected payoffs: player 1's is $[ap + c_h(1-p)] / [1 - pw]$ and player 2's is $[bp + d(1-p)] / [1 - pw]$. Because a strategy specifies players' moves in every period and the game's length is undetermined, infinitely many possible strategies exist. We are primarily interested in the seven strategies represented in Figure 2, which shows the expected payoff for each player for each strategy pair of interest.

Of particular interest is the strategy pair designated in the top left corner of figure 2), in which player 2 discriminates between cooperators who look and those who do not look, and player 1 cooperates without looking (CWOL). The following simple argument demonstrates that this strategy pair is a Nash Equilibrium (which means no player has an incentive to unilaterally deviate) whenever $\frac{a}{1-w} \geq c_l p + c_h(1-p)$. This condition has a natural interpretation: player 1's expected temptation from defection is less than the gains from an ongoing cooperative interaction. Thus, player 1 would lose from deviating, for example by looking, because this would end the lucrative ongoing relationship, reducing player 1's payoff from an expected $\frac{a}{1-w}$ to 0. Player 2 would also lose from deviating, by choosing to exit, because when player 1 is always cooperating, she expects $\frac{b}{1-w}$ from the relationship, and exiting yields 0.

One potential concern with this equilibrium is that, since player 2 is not worse off by not attending to looking, he might not do so. This turns out to not be the case. The intuition is that if there is even a small probability player 1 looks, player 2 is better off

attending to looking. In the appendix, we formalize this intuition by showing that this equilibrium is subgame perfect, a solution concept used to rule out these kinds of concerns in settings where there is a high degree of rationality [15].

We might wonder whether CWOL will emerge in a population of players who are not rational, but are evolving or learning their strategies. There are two reasons to suspect it will not emerge. First, CWOL may be susceptible to the invasion of mutants. In particular, the concern is that a player 2 mutant might arise that does not attend to whether player 1 looks, rendering looking irrelevant. This mutant's fitness is no higher or lower than the incumbent strategy, so may grow by drift. Subsequently, a player 1 mutant which looks and defects when the temptation is high would have a fitness advantage and proliferate. While the player 2 mutant would then start dying off since it would now have lower fitness than the incumbent, it is not clear that it would do so fast enough for the player 1 mutant to also die off, returning to the CWOL equilibrium. Second, CWOL may be stable but have such a small basin of attraction that it will never emerge. In particular, there are three other equilibria to consider, which can be seen in figure 2. The first is comprised of the strategy pair where player 1 always defects, and player 2 always exits (which we refer to as the ALLD equilibrium). It is a Nash Equilibrium for all parameter values. The second is the strategy pair where player 2 exits if player 1 defects, and player 1 cooperates with or without looking (we refer to this as the CWL equilibrium). It is a Nash equilibrium when $\frac{a}{1-w} \geq c_2$. In fact, there are many more Nash Equilibria, where the population mixes between different strategies.

Consequently, we employ computer simulations of the replicator dynamic to explore the evolutionary dynamics of the envelope game. The replicator dynamic is the standard model for evolutionary dynamics [16, 17, 18, 19], and also models learning dynamics such as reinforcement learning or prestige-biased imitation [20]. It describes strategies evolving

6

over time under the assumption that the rate of reproduction within each population is proportional to the fitness relative to that type's other strategies. Since replicator dynamics cannot be solved in closed form, they must analyzed using computer simulations seeded with randomly chosen strategy frequencies. This method also requires restricting the analysis to a few strategies. Therefore, we run our replicator analysis only on the restricted set of strategies represented in figure 2. While this set of strategies is restricted, it includes the potentially destabilizing mutants discussed above.

To investigate the frequencies in which different equilibria emerge, we randomly seed the strategy frequencies many times and record the frequency of each strategy after the population has stabilized. Many mixtures of strategies are behaviorally consistent with CWOL, for example, if a large enough fraction of player 2s exit when player 1s look then no player 1s will look, even if this fraction is less than 1. Thus, we classify population frequencies as being behaviorally equivalent to ALLD, CWL, and CWOL (see appendix for details). We plot the frequencies of each of these for different values of a in figure 3 (see appendix for analogous figures for other parameters). As can be seen, CWOL emerges often when it is a Nash Equilibrium.

Next, we identify the conditions under which people will be most likely to avoid looking and detect looking amongst others. For this, we interpret the conditions under which CWOL is the only cooperative equilibrium, because, as we show in the appendix, when one allows for mutations, the CWOL equilibrium no longer emerges in dynamic simulations when both CWOL and CWL are equilibria. Recall that CWL is a Nash equilibrium if and only if $\frac{a}{1-w} \geq c_2$. This equilibrium condition has a natural interpretation: in order to sustain CWL, the long term gains to player 1 from the ongoing relationship must suffice for player 1 to cooperate even when player 1 knows the temptation is high in the current period. Contrast this with the condition that determines if CWOL is an equilibrium:

7

$\frac{a}{1-w} \geq c_1 p + c_2(1-p)$. This equilibrium condition has a natural interpretation as well: in order to sustain CWOL, the long term gains to player 1 from the ongoing relationship must suffice for player 1 to cooperate when player 1 *expects* the temptation to *sometimes* be high. That is, not looking makes the expected–as opposed to realized–gains from defection relevant, in a sense smoothing the temptation to defection. Thus, the range where CWOL is an equilibrium and CWL is not, $c_1 p + c_2(1-p) \leq \frac{a}{1-w} < c_2$, has the following interpretation: the expected temptation is low but the maximal temptation is high. In the appendix, we confirm this result using our subgame perfections analysis. We also use dynamics to show that, CWOL increases relative to CWL when we increase the maximal temptation, but hold the mean temptation constant.

We identify a second condition under which people will be most likely to avoid and detect looking by relaxing the assumption $d > \frac{p}{1-p}b$. Then, in the region where $d \leq \frac{p}{1-p}b$, there is a fourth equilibrium. It is the strategy pair where player 2 always continues if player 1 cooperates when the temptation is low, and player 1 looks and cooperates only when the temptation is low (we refer to this as the ONLYL equilibrium). In contrast, CWOL is an equilibrium for all values of d. CWOL is thus the only cooperative equilibrium in the parameter region $d > \frac{p}{1-p}b$, which has the interpretation: defection is sufficiently harmful to player 2 such that player 2 prefers to avoid the interaction if player 1 only cooperates some of the time.

Note that CWOL is an equilibrium over a wider parameter region than both CWL and ONLYL, and thus that the ability to avoid looking and to detect whether others look increases the parameter space over which cooperation is feasible. To see this, consider removing player 1's strategy which consists of not looking. Alternatively, consider removing player 2's strategy where he conditions his behavior on whether player 1 has looked. In either case, cooperation is only sustained in a Nash Equilibrium if $\frac{a}{1-w} \geq ap + c_2(1-p)$ or

8

$d > \frac{p}{1-p}b$. These are the same condition needed for CWL and ONLYL, respectively, and, as discussed above, these ranges is strictly smaller than the range of values over which CWOL is an equilibrium.

To summarize, CWOL can occur in equilibrium, and this equilibrium is subgame perfect, is stable, and has a sizable basin of attraction in the replicator dynamics. Moreover, we expect player 2 to prefer interacting with player 1s who do not look and player 1s to actively avoid looking when defection is harmful and the temptation to defect is usually small but sometimes huge. Finally, under these conditions, cooperation can be sustained only if player 1 can avoid looking and player 2 can observe whether player 1 looks.

While we modeled player 1 looking at the temptation to defect, we could analogously have modeled player 1 looking at the benefits to cooperating. This can be interpreted as looking to see if anyone is watching, asking what one will get, or calculating the value of the ongoing relationship. In this and some similar cases, the modifications to the model would be straightforward and analysis would be equivalent.

We now apply the model to shed light on a number of interesting phenomena, including why we dislike "flip-flopping" politicians and respect principled people more generally, why people cooperate intuitively, why people feel disgust when considering taboo trade-offs, and why people fall in love.

We trust candidates for political office whose policies are the result of their convictions and are consistent over time, and distrust those whose policies are carefully constructed in consultation with their pollsters and who "flip-flop" in response to public opinion, as caricatured by the infamous 2004 Republican presidential campaign television ad showing John Kerry wind-surfing, and tacking from one direction to another. At first glance, this seems irrational: one would think it a virtue for a politician to flexibly respond to public opinion once in office. This logic is illustrated by John Maynard Keynes' famous quote:

9

"When the facts change, I change my mind. What do you do, sir?" However, if policies based on conviction or consistency over time are signals that a candidate does not look at the benefits of cooperation, this is an indication that the candidate can be trusted in situations where making the decision that is right for his or her constituency comes at a large political cost. Consistent with this, opponents take every opportunity to paint candidates as flip-floppers who cannot be trusted [21]. This argument generalizes outside of politics to why we respect people who are "principled" over those who are "strategic".

People intuitively cooperate. That is, when people make decisions rapidly, they are more likely to cooperate than if they have time to deliberate. Additionally, people who cooperate make quicker decisions than those who defect [22]. The Social Heuristics Hypothesis offers one explanation for this phenomena: we adopt heuristics to avoid incurring cognitive costs associated with deliberation [23, 24, 25]. In a world with repeated interactions, it is usually worthwhile to cooperate, so individuals may adopt heuristics such as "always cooperate" or "cooperate as long a situation is not a business interaction." These same individuals, when serving as laboratory subjects, may apply these heuristics and cooperate even when it is not worthwhile to do so [22, 26].

Our model offers the following alternative explanation for intuitive cooperation. Intuitive cooperation may serve to reduce responsiveness to realized costs. Thus, others should trust intuitive cooperators more than deliberate cooperators. Since intuitive cooperators are more trustworthy, this may lead people to evolve or learn to become intuitive cooperators. Note that this explanation suggests that intuitive cooperation may be an optimal response to others' ability to detect deliberation, and not, as a heuristic is, an attempt to avoid cognitive costs. For this explanation to be sensible, it must be the case that whether or not a decision is made intuitively or deliberatively is detectable. In fact, it is: deliberative decision making leads to slower reaction time, as well as increased pupil size

10

and heart rate [27], and sometimes blushing or stammering [28]. In contrast to the Social Heuristics Hypothesis, our model also predicts that decisions related to cooperation are more likely to be intuitive than other decisions that are similarly usually worthwhile, and that intuitive cooperators are trusted more than reflective cooperators. To our knowledge, neither of these predictions has been tested yet.

People dislike considering trade-offs related to "sacred values" [29]. Sacred values are values such as love, liberty, honor, justice, or life, that people treat "as possessing transcendental significance that precludes comparisons, trade-offs, or indeed any mingling with secular values" [29]. While there is variation in what societies consider sacred, virtually all societies have a concept of sacredness [29]. Sacred values are so strongly imbued in us that we do not find them puzzling prima fascia, yet their existence and origin remains poorly understood. What makes us treat some values as sacred and what differentiates these values from secular values like free time or money that we more readily trade off?

Our model provides one possible explanation. People who calculate costs of trading off against sacred values are less trustworthy when it comes to safeguarding these values than people who consider them sacred and would never calculate the costs of trading off against them. Responding with disgust to these "taboo trade-offs" may be one way to prevent us from interacting with people who make such trade offs and hence are less trustworthy, and may also be a way to signal to others that we ourselves would not consider, and therefore make, such trade-offs. Consistent with CWOL, it is taboo to consider the trade-off even if one ultimately makes the right choice, and the longer the trade-off is considered for, the harsher the judgement by observers [29]. As with intuitive cooperation, if people who refuse to consider taboo trade-offs are seen as more trustworthy, this may lead people to evolve or learn to abide by these taboos. If CWOL indeed underlies

11

the phenomena of taboo trade-offs, then it provides a novel prediction: taboo trade-offs will prevail precisely in situations where there is large but infrequent temptation to defect and defection is harmful, such as selling a child, betraying a country, or sleeping with someone for a million dollars. It remains to be shown that taboo trade-offs demonstrate this characteristic. It also provides an important policy prescription regarding policies forbidding taboo trade-offs, for example, the ban on euthanasia: such policies are socially suboptimal, since the benefits of cooperating without looking accrue to the individuals who advocate them, but the costs are borne by society.

Finally, our model offers an explanation for emotions such as love, and can thus be thought of as a formalization of Frank (1988). Love has the property that we behave altruistically towards our partners regardless of what temptations arise [30], as illustrated by the wedding vow, "for better or for worse, for richer, for poorer, in sickness and in health." For example, love causes individuals to ignore other potential mates, even if those mates are better than one's current mate, as Shakespeare's Juliet did when her love for Romeo led her to rebuff the advances of the otherwise-more-suitable Paris.

Why does love have this property? Our model suggests that those in love will more often be chosen as long-term partners. If this argument in fact underlies love, then it is crucial that love is observable, and that it necessarily cannot be displayed while still attending to costs. There is evidence consistent with this: related emotions are observable [31], cannot be faked [32], and are relied upon by partners when choosing whether to cooperate [33]. There is also reason to believe love and related emotions would be hard to fake, given their autonomic origins, and the costs of placing their activation under conscious control [30, 28]. However, it remains to be shown that love in particular has these attributes, and that it cannot be displayed while attending to costs.

Our formalism adds novel insights to Frank's argument about love. First, our model

clarifies that falling in or out of love depends on the distribution of temptations, but not their immediate realizations. This suggests people will fall out of love when there is a permanent change in alternative mating opportunities or relationship costs, but not a one-off temptation. For example, a man may fall out of love with his wife after becoming unexpectedly successful, as many more women will be now be interested in him than he anticipated when they first met. Second, the model clarifies that love comes with a cost–the cost of ignored temptations–and suggests that this cost must be compensated with commensurate investment in the relationship. Only sometimes is it worthwhile for the recipient of love to compensate a suitor, which explains why people actively avoid the strong affections of those with whom they do not wish to have long-term relationships. Third, our model clarifies why mere discussions of the costs and benefits of a relationship or a break-up, for example, suggesting a prenuptial agreement, damage the relationship. Such discussions indicate that one is looking at the costs of the relationship and cast doubt on one's commitment.

These arguments extend to anger. Anger can be thought of as "punishing without looking". It prevents people from looking at the costs of inflicting harm on others after a transgression, thereby deterring future transgressions.

Acknowledgments

References

[1] R. L. Trivers, *Quarterly review of biology* pp. 35–57 (1971).

[2] J. W. Friedman, *The Review of Economic Studies* **38**, 1 (1971).

[3] R. Axelrod, W. D. Hamilton, *Science* **211**, 1390 (1981).

[4] D. Fudenberg, E. Maskin, *Econometrica: Journal of the Econometric Society* pp. 533–554 (1986).

[5] D. Fudenberg, E. Maskin, *The American Economic Review* **80**, 274 (1990).

[6] K. G. Binmore, L. Samuelson, *Journal of economic theory* **57**, 278 (1992).

[7] M. A. Nowak, K. Sigmund, *Nature* **355**, 250 (1992).

[8] M. Nowak, K. Sigmund, *et al.*, *Nature* **364**, 56 (1993).

[9] R. J. Aumann, L. S. Shapley, *Long-term competitiona game-theoretic analysis* (Springer, 1994).

[10] M. A. Nowak, K. Sigmund, *Nature* **393**, 573 (1998).

[11] M. A. Nowak, K. Sigmund, *Nature* **437**, 1291 (2005).

[12] M. A. Nowak, *science* **314**, 1560 (2006).

[13] H. Ohtsuki, Y. Iwasa, *Journal of Theoretical Biology* **239**, 435 (2006).

[14] K. Sigmund, *The calculus of selfishness* (Princeton University Press, 2010).

[15] M. J. Osborne, *An introduction to game theory* (Oxford University Press, USA, 2003).

[16] P. D. Taylor, L. B. Jonker, *Mathematical biosciences* **40**, 145 (1978).

[17] J. W. Weibull, *Evolutionary game theory* (MIT press, 1997).

[18] J. Hofbauer, K. Sigmund, *Evolutionary games and population dynamics* (Cambridge University Press, 1998).

[19] M. A. Nowak, *Evolutionary dynamics: exploring the equations of life* (Harvard University Press, 2006).

[20] D. A. Fudenberg, *The theory of learning in games*, vol. 2 (MIT press, 1998).

[21] J. Markon, Obama fires up crowd in virginia with 'romnesia' speech, *The Washington Post*, October 19, 2012, Available: `http://articles.washingtonpost.com/2012-10-19/politics/35499645_1_romnesia-obama-fires-economy-with-higher-taxes` [Last accessed: August 11, 2013].

[22] D. G. Rand, J. D. Greene, M. A. Nowak, *Nature* **489**, 427 (2012).

[23] H. A. Simon, *The quarterly journal of economics* **69**, 99 (1955).

[24] A. Tversky, D. Kahneman, *Science* **185**, 1124 (1974).

[25] D. Kahneman, P. Slovic, A. Tversky, *Judgment under uncertainty: Heuristics and biases* (Cambridge University Press, 1982).

[26] D. Rand, *et al.*, *Available at SSRN: h ttp://ssrn. com/abstract* **2222683** (2013).

[27] D. Kahneman, B. Tursky, D. Shapiro, A. Crider, *Journal of Experimental Psychology* **79**, 164 (1969).

[28] S. Pinker, *NY: Norton* (1997).

[29] P. E. Tetlock, *Trends in cognitive sciences* **7**, 320 (2003).

[30] R. H. Frank, *Passions within reason: The strategic role of the emotions.* (WW Norton & Co, 1988).

[31] P. Ekman, E. R. Sorenson, W. V. Friesen, *Science* **164**, 86 (1969).

[32] P. Ekman, R. J. Davidson, W. V. Friesen, *Journal of personality and social psychology* **58**, 342 (1990).

[33] L. I. Reed, K. N. Zeglen, K. L. Schmidt, *Evolution and Human Behavior* **33**, 200 (2012).
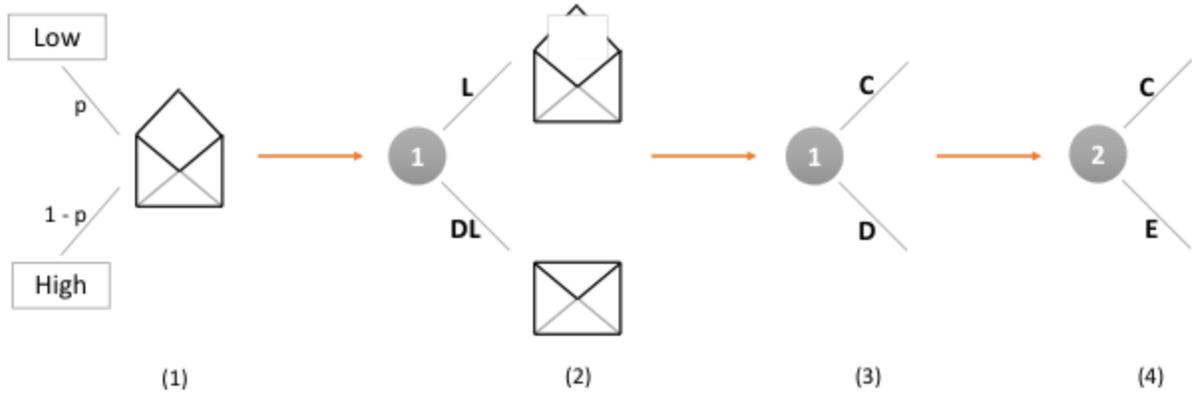
FIGURE 1: THE ENVELOPE GAME

FIGURE 2: PAYOFFS FOR A RESTRICTED SET OF STRATEGIES IN THE ENVELOPE GAME

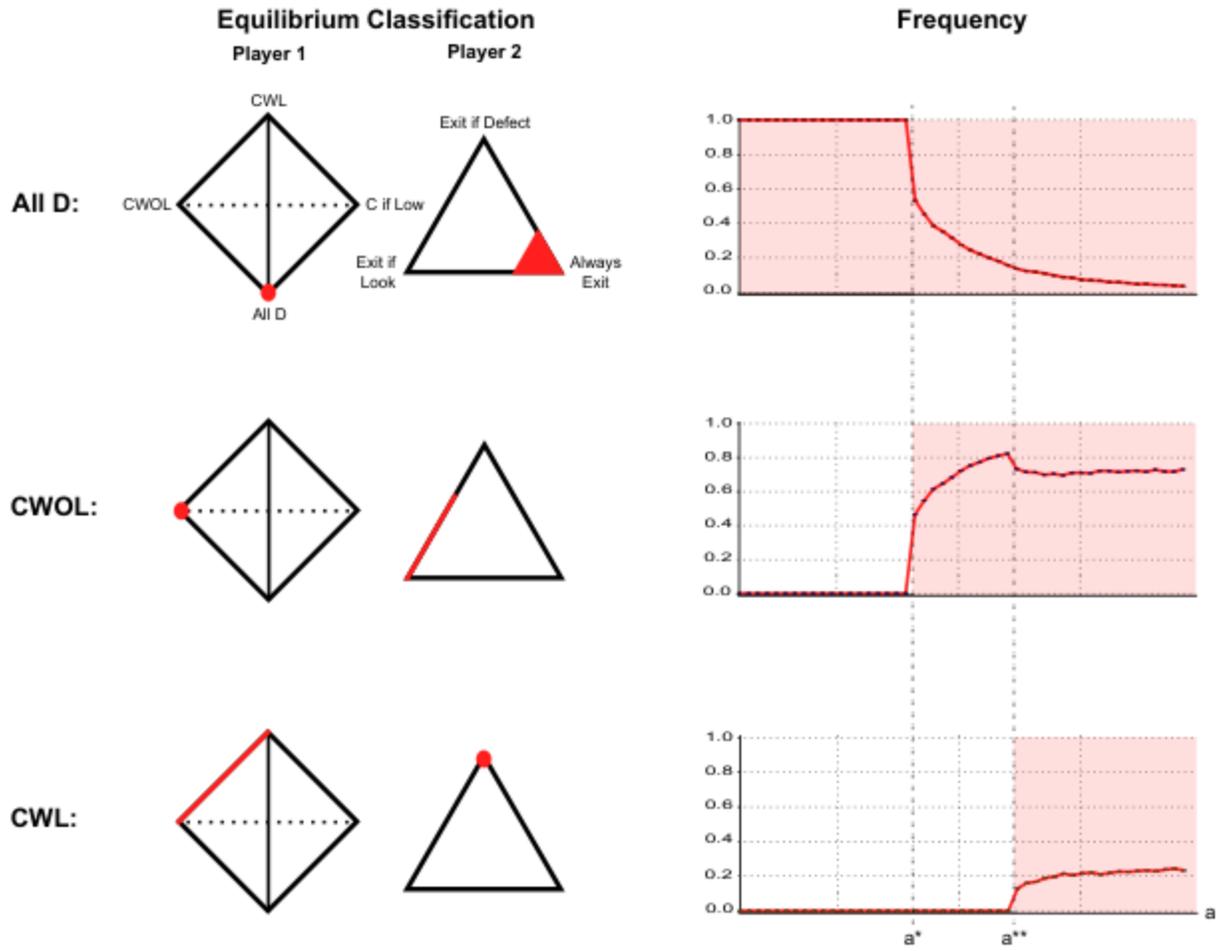| | Player 2 | | |
Player 1	Continue if Player 1 Cooperates Without Looking	Continue if Player 1 Cooperates	Always Exit
Cooperate Without Looking	$\frac{a}{1-w}, \frac{b}{1-w}$	$\frac{a}{1-w}, \frac{b}{1-w}$	a, b
Cooperate With Looking	a, b	$\frac{a}{1-w}, \frac{b}{1-w}$	a, b
Look and Cooperate Only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

18

Figure Legends

Figure 1: The Envelope Game

We model non-strategic cooperative behavior using what we call the envelope game. (a) Column 1: The game begins when the temptation to defect is randomly chosen, as indicated by a notice randomly being placed in the envelope. The temptation to defect is low with probability p and high with probability $1 - p$. Column 2: Then, player 1 chooses whether to look (open the envelope) or not. Column 3: Player 1 then chooses whether to cooperate or defect. Player 1 may only condition her action on the realized temptation determined in column 1 if she looked. Each time player 1 cooperates, then, regardless of whether player 1 looked, player 1 gets $a > 0$ and player 2 gets $b > 0$. Each time player 1 defects, her payoffs depend on whether defection was tempting. If it was not tempting, player 1 gets $c_l > a$ and if it was tempting, player 1 gets $c_h > c_l$. In either case, each time player 1 defects, player 2 gets $d < 0$. Column 4: Player 2, having observed both of player 1's choices, chooses whether to continue or exit. If player 2 continues, with probability w, all previous steps are repeated, potentially indefinitely.

Figure 2: Payoffs for a Restricted Set of Strategies in the Envelope Game

This table presents the payoffs for the restricted set of strategies used for the replicator analysis. Player 1's strategies are presented in separate rows, and player 2's strategies are presented in columns. The payoffs presented in the intersection of a given row and column are those that the players receive if they play the corresponding strategies. For example, consider what happens if player 1 looks and cooperates only if the temptation is low (penultimate row) and player 2 repeats continues when player 1 cooperates (middle column). Then, player 1's expected payoff is $[ap + c_h(1 - p)][1 - pw]$ (the first entry in the corresponding cell) and player 2's is $[bp + d(1 - p)][1 - pw]$ (the second entry in the

20

same cell; for details of calculations leading to payoffs, see appendix). This calculation is the result of the following logic: each time the temptation is low, player 1 cooperates and gets a, player 2 gets b, and the game continues with probability w until the first time the temptation is high. We refer to the strategy where player 1 cooperates without looking (top row) as CWOL. We also refer to the strategy pair where player 1 CWOLs and player 2 continues if player 1 CWOLs (first column) as CWOL. We refer to the strategy pairs where player 1 cooperates with or without looking and player 2 continues if player 1 cooperates (first and second row, and middle middle column) as CWL. We refer to the strategy pair where player 1 always defects and player 2 always exits (bottom row and rightmost column) as ALLD. ALLD is always an equilibrium of the envelope game. CWOL is an equilibrium if $a/(1-w) > c_l p + c_h(1-p)$. CWL is an equilibrium if $a/(1-w) > c_h$. This region is a subset of the region for which CWOL is an equilibrium.

Figure 3: Learning Dynamics of the Envelope Game

We apply the replicator dynamic to the envelope game restricted to the strategies represented in figure 2. The replicator dynamic describes strategies evolving over time under the assumption that the rate of reproduction within each population is proportional to the fitness relative to that type's other strategies. The replicator dynamic also models learning dynamics such as reinforcement learning or prestige-biased imitation. We run 1000 time series with randomly seeded strategy frequencies for a range of values of a, and record the frequency with which they stabilize in one of the strategy pairs identified in figure 2, or in a behaviorally equivalent equilibrium, as presented in the simplexes. We vary the value of a along the x-axis. The y-axis represents frequencies, and each colored line presents the frequency of the strategy pair. The parameter region where the strategy pair is supported as an equilibrium is shaded in light red. CWOL is supported as an

21

equilibrium in the region $a \geq a^* = (1 - w) \cdot [c_l p + c_h (1 - p)]$, and emerges often in this parameter region. CWL is supported as an equilibrium in a strictly smaller parameter region, $a \geq a^* = (1 - w) \cdot c_h$. Neither CWOL nor CWL emerge in the parameter regions where they are not supported in equilibrium.