Supplementary Information for Cooperate Without Looking

Moshe Hoffman, Erez Yoeli, Martin A. Nowak

Program for Evolutionary Dynamics,
Harvard University, Cambridge, MA 02138

December 19, 2014

Contents

List of Figures

A Calculation of Payoffs in the Envelope Game

In this section, we describe how the payoffs in Fig. 2 of the manuscript are calculated. For ease of exposition, we present the payoffs again below. We color-code them to assist with our description.

FIGURE S1: PAYOFFS FOR A RESTRICTED SET OF STRATEGIES IN THE ENVELOPE GAME

	Player 2		
Player 1	End if Player 1 Looks	End if Player 1 Defects	Always End
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	$c_l p + c_h(1-p), d$	$c_l p + c_h(1-p), d$	$c_l p + c_h(1-p), d$

First, consider the strategy pairs whose payoffs are highlighted in red. For these strategy pairs, the game only lasts one round. For strategy pairs where player 1 cooperates, the payoffs in this one round are a, b. For strategy pairs where player 1 cooperates if the temptation is low, players' expected payoffs are $ap + (1-p)c_h, bp + (1-p)d$. For strategy pairs where player 1 defects, players' expected payoffs are $pc_l + (1-p)c_h, pb + (1-p)d$.

Next, consider the strategy pairs whose payoffs are highlighted in blue. In all of these strategies, on the equilibrium path, player 1 always cooperates and player 2 always continues. At each stage, players get a, b, and the game continues with probability w, so players get $a + wa + w^2a + w^3a + \ldots, b + wb + w^2b + w^3b + \ldots$. Solving this geometric series yields payoffs $\frac{1}{1-w}a$, $\frac{1}{1-w}b$.

Finally, consider the last remaining strategy pair, where player 1 cooperates when the temptation is low and player 2 ends iff player 1 defects when the temptation is low. In this case, players' expected payoff in the first round is, $pa + (1-p)c_h, pb + (1-p)d$. Players continue to the second round only if the temptation was low in the first round. The likelihood of continuing to the second round is therefore pw. Players' expected payoffs are therefore, $pa + (1-p)c_h + pw[pa + (1-p)c_h] +$

$(pw)^2[pa + (1-p)c_h] + \ldots, pb + (1-p)d + pw[pb + (1-p)d] + (pw)^2[pb + (1-p)d] + \ldots$. Solving this geometric series yields payoffs $\frac{pa+(1-p)c_h}{1-pw}, \frac{pb+(1-p)d}{1-pw}$.

B Proofs

B.1 Nash Equilibria of the Restricted Envelope Game

We begin by considering only the seven strategies employed in the manuscript. The proofs in this section follow from direct inspection of the payoff matrix presented in Fig. 2 (and Fig. S1).

Claim 1. *The strategy pair (always defect, always end) is always a Nash equilibrium of the restricted envelope game.*

Proof. We first show that player 1 cannot benefit by deviating to another strategy when player 2 always ends. Player 1's payoff if she does not deviate is $c_l p + c_h(1-p)$. If player 1 deviates to look and cooperate only when the temptation is low, her payoff is $ap + c_h(1-p)$. Since $a < c_l$, this is lower than her payoff if she does not deviate. If player 1 deviates to either CWOL or CWL, her payoff is a, which, since $a < c_h$, is even lower. Thus, she cannot benefit by deviating.

We next show that player 2 cannot benefit by deviating to another strategy when player 1 always defects. For all three of player 2's strategies, her payoff is the same, so she cannot benefit from deviating. □

Claim 2. *Iff $\frac{a}{1-w} \geq c_l p + c_h(1-p)$, the strategy pair (cooperate without looking, end if player 1 looks) is a Nash equilibrium of the restricted envelope game.*

Proof. Assume $\frac{a}{1-w} \geq c_l p + c_h(1-p)$. We first show that player 1 cannot benefit by deviating to another strategy. Player 1's payoff if she does not deviate is $\frac{a}{1-w}$. Of the three strategies player 1 could deviate to, always exit has the highest payoff, $c_l p + c_h(1-p)$. By assumption, this payoff is less than or equal to $\frac{a}{1-w}$, so player 1 does not benefit by deviating.

We next show that player 2 cannot benefit by deviating to another strategy. Player 2's payoff if she does not deviate is $\frac{b}{1-w}$. If she deviates to end if player 1 defects, she gets the same payoff. If she deviates to always end, she gets b, which is worse. So player 2 does not benefit from deviating.

Now, assume $\frac{a}{1-w} < c_l p + c_h(1-p)$. Then, player 1's payoff from always exit is higher than the payoff from cooperate without looking, and she deviates. □

Claim 3. *Iff $\frac{a}{1-w} \geq c_h$, the strategy pair (cooperate with looking, end if player 1 defects) is a Nash equilibrium of the restricted envelope game.*

Proof. Assume $\frac{a}{1-w} \geq c_h$. We first show that player 1 cannot benefit by deviating to another strategy. Player 1's payoff if she does not deviate is $\frac{a}{1-w}$. Player 1 cannot benefit by deviating to cooperate without looking, since this yields the same payoff. If she deviates to look and cooperate only when the temptation is low, she gets $\frac{ap+c_h(1-p)}{1-pw}$. Some arithmetic shows that $\frac{ap+c_h(1-p)}{1-pw} \leq \frac{a}{1-w}$ iff $c_h \leq \frac{a}{1-w}$, which is true by assumption. Finally, if she deviates to always defect, she gets $c_l p + c_h(1-p) < c_h$, and does not benefit.

We next show that player 2 cannot benefit by deviating to another strategy. Player 2's payoff if she does not deviate is $\frac{b}{1-w}$. If she deviates to end if player 1 looks or to always end, she gets b, which is worse. So, player 2 does not benefit from deviating.

Now, assume $\frac{a}{1-w} < c_h$. Then, player 1's payoff from switching to always exit when the temptation is high is greater than her current payoff, so she deviates. □

B.2 Considering a Richer Strategy Set

We now consider all strategies of the Envelope Game. Because we restricted the strategy space in the manuscript, it was not necessary to fully define the strategies considered. We do so now.

In the Envelope Game, Player 1's pure (or deterministic) strategy specifies (a) whether she will look as a function of her past behavior and (b) whether she will cooperate as a function of her past behavior and the temptation, if she has just looked. Similarly, a strategy for player 2 specifies whether she will end as a function of player 1's past behavior. For example, the strategy "cooperate without looking in odd periods and defect without looking in even periods" is a legitimate strategy for player 1, even though such a strategy is probably unrealistic. Another example of a legitimate strategy for player 1 is, "cooperate without looking iff player 1 has always cooperated without looking in the past."

We formally define the strategy pairs ALLD, CWOL, and CWL discussed in the manuscript as follows:

- ALLD: (look and defect regardless of how player 1 has played in the past, end regardless of how player 1 has played in the past)
- CWOL: (cooperate without looking regardless of how player 1 has played in the past, end if player 1 looked or defected in the last round)
- CWL: (look and cooperate regardless of how player 1 has played in the past, end iff player 1 defected in the last round)

Claim 4. *ALLD is always a Nash equilibrium of the Envelope Game.*

Proof. We first consider whether player 1 can benefit by cooperating in the first round. If player 2 always ends, player 1 gets no benefit from paying the cost of cooperating in this round, so he does not deviate. Likewise, if player 1 always defects, player 2 gets negative payoffs from continuing. Because the game ends after the first round, any other unilateral deviation will not influence payoffs, and hence players cannot benefit from this deviation. □

Claim 5. *Iff* $\frac{a}{1-w} \geq c_l p + c_h(1-p)$*, CWOL is a Nash equilibrium of the Envelope Game.*

Proof. Suppose $\frac{a}{1-w} < c_l p + c_h(1-p)$ but player 1 cooperates without looking and player 2 ends iff player 1 looks or defects. Then player 1 can deviate by defecting, in which case she gets $c_l p + c_h(1-p)$ which is higher than $a + aw/(1-w) = \frac{a}{1-w}$.

Suppose $\frac{a}{1-w} \geq c_l p + c_h(1-p)$. Player 1 may deviate at any point by either looking or defecting (or both). The most beneficial deviation involves defecting, yielding an expected payoff of $c_l p + c_h(1-p)$, which is assumed to be lower than her expected payoff from playing the equilibrium strategy of $\frac{a}{1-w}$. Player 2 may deviate at any point by ending. She expects $\frac{b}{1-w}$ in the future if she continues and 0 in the future if she ends. By assumption, the former is larger, so she is better off not deviating. □

Claim 6. *Iff* $\frac{a}{1-w} \geq c_h$*, CWL is a Nash equilibrium of the Envelope Game.*

Proof. Suppose $\frac{a}{1-w} < c_h$ but player 1 cooperates and player 2 ends iff player 1 defects. Then player 1 can deviate by defecting when the temptation is high, in which case she gets c_h, which, by assumption, is higher than her payoff from continuing to cooperate, $\frac{a}{1-w}$.

Suppose $\frac{a}{1-w} \geq c_h$. Player 1 may deviate at any point by not looking or by defecting. Not looking yields no benefit since player 2 will continue regardless of whether player 1 looks. Since player 1 has looked, she can condition deviation on the realized temptation. In particular, the best she can do is to defect whenever there is a high temptation. In this case she would get a payoff of c_h, which is assumed to be no greater than her expected payoff from playing the equilibrium strategy of $\frac{a}{1-w}$. Player 2 may deviate at any point by ending. She expects $\frac{b}{1-w}$ in the future if she continues and 0 in the future if she ends. By assumption, the former is larger, so she is better off not deviating. □

B.3 Subgame Perfection

The above strategy pairs, while Nash, have the following pitfall. Consider CWOL. Suppose player 2 were to find herself in the circumstance (i.e. "subgame,") where player 1 had defected. Player 2 could do better by continuing instead of exiting because player 1 would not be expected to defect again. This strategy pair thus specifies that player 2 behave suboptimally, albeit in a subgame that does not happen according to the strategy pair.

Subgame perfection rules out such strange behavior by requiring players to be playing a Nash equilibrium in every subgame. Why would we be concerned with Nash equilibria that are not subgame perfect? First, if players are playing a Nash equilibrium that is not subgame perfect, then they must believe that the other would behave suboptimally if the other finds herself in a particular subgame. For instance, if player 1 were to defect, she would expect player 2 to exit, even though player 2 would be better off continuing. Thus, if players are not only rational, but also know that others are rational, and perhaps even know that others know they are rational, etc.–this assumption is known as "common knowledge of rationality" [1]–then Nash that are not subgame perfect are suspect. Second, even if players are not rationally thinking through these off-equilibrium-path contingencies, a second justification for subgame perfection are "trembles". If, every once in a while, players tremble, i.e. mistakenly choose the wrong action, then every subgame is reached with some, albeit small, probability. Then, only subgame perfect equilibria are Nash equilibria of this new game with trembles [2]. For instance, every once in a while, player 1 would accidentally choose to defect. A player 2 who would exit after such a mistake could do better by deviating to continue because she can expect continued cooperation thereafter.

While the strategies in section B.2 are not subgame perfect, we now specify strategies that are subgame perfect and are "outcome equivalent" to the aforementioned strategies, meaning that if both players play according to the strategies prescribed below, the resulting behavior will be the same as if they had both played the strategies in section B.2.

- ALLD: (look and defect regardless of how player 1 has played in the past, end regardless of how player 1 has played in the past)
- CWOL: (cooperate without looking iff player 1 has never looked or defected and defect otherwise, end iff player 1 has ever looked or defected)
- CWL: (look and cooperate iff player 1 has never defected and defect otherwise, end iff player 1 has ever defected)

Claim 7. *ALLD is always a subgame perfect equilibrium of the Envelope Game.*

Proof. We first consider whether player 1 can benefit by cooperating in the first round. If player 2 always ends, player 1 gets no benefit from paying the cost of cooperating in this round, so he does not deviate. Likewise, if player 1 always defects, player 2 gets negative payoffs from continuing. Because the game ends after the first round, any other unilateral deviation will not influence payoffs, and hence players cannot benefit from this deviation. □

Claim 8. *Iff* $\frac{a}{1-w} \geq c_l p + c_h(1-p)$*, CWOL is a subgame perfect equilibrium of the Envelope Game.*

Proof. Suppose $\frac{a}{1-w} < c_l p + c_h(1-p)$ but player 1 cooperates without looking and player 2 ends iff player 1 looks or defects. Then player 1 can deviate by defecting, in which case she gets $c_l p + c_h(1-p)$ which is higher than $a + aw/(1-w) = \frac{a}{1-w}$.

Suppose $\frac{a}{1-w} \geq c_l p + c_h(1-p)$. We consider all four one-shot deviations (this includes deviations off the equilibrium path). First, player 1 may deviate in the first iteration. Since any such deviation will cause player 2 to end, the most beneficial deviation involves defecting in both periods, yielding $c_l p + c_h(1-p)$, which is assumed to be lower than her expected payoff from playing the equilibrium strategy of $\frac{a}{1-w}$. Now suppose player 2 has always observed that player 1 cooperates without looking. Then she expects $\frac{b}{1-w}$ in the future if she continues and 0 in the future if she ends. By assumption, the former is larger, so she is better off not deviating. Now suppose player 1 finds herself in the history where she has deviated in the past and player 2 has nevertheless continued. Player 1 prefers to defect now, since player 2 will end regardless, and cooperation is costly by assumption. Finally, suppose player 2 finds herself in the history where player 1 has deviated. If player 2 ends as prescribed, she gets 0 in the future, while if she continues, player 1 will defect, and player 2's expected payoff is dw. This is negative by assumption. □

The following, more general result, follows from claim 8.

Corollary 1. *Iff $\frac{a}{1-w} \geq c_l p + c_h(1-p)$, there exists a strategy pair that is subgame perfect and is outcome equivalent to CWOL.*

Proof. We have shown that if $\frac{a}{1-w} \geq c_l p + c_h(1-p)$, there exists a strategy pair that is subgame perfect and is outcome equivalent to CWOL in claim 8.

To show that if there exists a strategy pair that is subgame perfect and is outcome equivalent to CWOL, then $\frac{a}{1-w} \geq c_l p + c_h(1-p)$, we demonstrate that any strategy pair that is outcome equivalent to CWOL is not subgame perfect when the inequality does not hold. In particular, if the strategy pair is outcome equivalent to CWOL then player 1 must be cooperating without looking. In a given round, she can do better in expectation by deviating and defecting. □

Claim 9. *Iff $\frac{a}{1-w} \geq c_h$, CWL is a subgame perfect equilibrium of the Envelope Game.*

Proof. Suppose $\frac{a}{1-w} < c_h$. Then player 1 can deviate by observing the temptation and choosing to defect if the temptation is high. When this happens, she will get c_h that period and 0 ever after. She foregoes $\frac{a}{1-w}$, which, by assumption is less than c_h. Therefore, she is better off with this deviation.

Suppose $\frac{a}{1-w} \geq c_h$. We consider all four one-shot deviations. First, player 1 may deviate by defecting when she observes a high temptation. This will cause player 2 to end. Player 2's payoffs will thus be c_h in this period and 0 in the future instead of $\frac{a}{1-w}$, which, by assumption, is greater. Player 1 will be even worse off if she deviates when she observes a low temptation. Now suppose player 2 has always observed that player 1 cooperates. Then she expects $\frac{b}{1-w}$ in the future if she continues, and 0 if she ends. By assumption, the former is larger, so she is better off not deviating. Now suppose player 1 finds herself in the history where she has deviated in the past and player 2 has nevertheless continued. Player 1 prefers to defect now, since player 2 will end regardless, and cooperation is costly. Finally, suppose player 2 finds herself in the history where player 1 has ever defected. If player 2 ends as prescribed, she gets 0 in the future, but if she continues, player 1 will defect the next round, and player 2's expected payoff is dw. This is negative by assumption. □

The following, more general result, follows from claim 9.

Corollary 2. *Iff $\frac{a}{1-w} \geq c_h$, there exists a strategy pair that is subgame perfect and is outcome equivalent to CWL.*

Proof. We have shown that if $\frac{a}{1-w} \geq c_h$, there exists a strategy pair that is subgame perfect and is outcome equivalent to CWL in claim 9.

To show that if there exists a strategy pair that is subgame perfect and is outcome equivalent to CWL, then $\frac{a}{1-w} \geq c_h$, we demonstrate that any strategy pair that is outcome equivalent to CWL is not subgame perfect when the inequality does not hold. In particular, if the strategy pair is outcome equivalent to CWL then player 1 must be cooperating with looking. In a round with high temptation, she can do better by deviating and defecting. □

B.4 Classifying All Stationary Subgame Perfect Equilibria

Notice that there are many other subgame perfect equilibria. For instance, consider the strategy pair (cooperate without looking on odd periods and defect without looking on even periods so long as player 1 has always cooperated without looking on odd periods and defected without looking on even periods; look and defect otherwise, continue iff player 1 has always cooperated without looking on odd periods and defected without looking on even periods). This pair is subgame perfect under different conditions from those given in the manuscript. This may draw into question our statement that looking matters under these conditions. We leave for future research a more thorough investigation of such equilibria, but note that such equilibria are non-stationary, and hence are arguably less likely to evolve or be learned, as briefly mentioned above. We can make some additional, stronger, claims by restricting the analysis to pure strategy pairs which yield stationary outcomes, which we define as outcomes that are the same in every round in which the temptation is the same [3].

First note all strategy pairs which yield stationary outcomes yield one of the following six outcomes for player 1:

1. always defect with or without looking (outcome equivalent to ALLD)
2. cooperate without looking (outcome equivalent to CWOL)
3. cooperate with looking (outcome equivalent to CWL)
4. look and cooperate iff the temptation is low
5. look and cooperate iff the temptation is high

Note that outcomes 4 and 5 cannot be observed as part of a subgame perfect equilibrium because $pb + (1-p)d < 0$ and $p > 1/2$. This argument, combined with claims 7 - 9, implies the following three claims.

Claim 10. *Iff* $\frac{a}{1-w} < c_l p + c_h(1-p)$, *the only stationary subgame perfect equilibria are outcome equivalent to ALLD.*

Claim 11. *Iff* $c_h > \frac{a}{1-w} \geq c_l p + c_h(1-p)$, *the only stationary subgame perfect equilibria are outcome equivalent to CWOL or ALLD.*

Claim 12. *Iff $\frac{a}{1-w} \geq c_h$, the only stationary subgame perfect equilibria are outcome equivalent to ALLD, CWOL, or CWL.*

These claims provide further support for our claim in the manuscript that $c_h > \frac{a}{1-w} \geq c_l p + c_h(1-p)$ is the region where we should be most likely to discover CWOL.

C Evolutionary Dynamics of the Envelope Game

In this section, we clarify how outcomes of the simulation are classified to correspond with the three Nash equilibria of the Envelope Game identified in the manuscript: ALLD, CWOL, and CWL. We also provide details of the parameters and methods used in the simulations presented in Figure 3 of the manuscript. Matlab code is available from the authors upon request.

C.1 Classification of Outcomes

We classify population frequencies of the restricted envelope game that correspond to the Nash equilibria ALLD, CWOL, and CWL. Notice that the resulting categories are not exhaustive. Indeed, the majority of states are not categorized into one of the three categories.

Note that we classify as an equilibrium any state that is within a small tolerance (0.001) of the conditions stated below.

CWOL: We categorize as CWOL all the states in which all player 1s are playing CWOL and at least some player 2s are attending to looking.

Let β_{CWOL} represent the proportion of player 1s who cooperate without looking, δ_{EID} be the proportion of player 2s who end iff player 1 defects, and δ_{EIL} be the proportion of player 2s who end iff player 1 looks or defects. We classify as CWOL any population state in which $\beta_{CWOL} = 1$, $\delta_{EID} + \delta_{EIL} = 1$ and

$$\frac{a}{1-w} \geq \frac{pa + (1-p)c_h}{1 - pw\delta_{EID}}.$$

This condition's interpretation is that the payoffs to player 1 from cooperating without looking are greater than the payoffs from cooperating only when the temptation is low given the proportion of player 2's who are attending to looking. This condition ensures that player 2s are attending to looking sufficiently often that player 1s do not benefit from deviating to other strategies, namely the binding strategy of looking and defecting when the temptation is high.

CWL: We categorize as CWL all the states where all player 1s cooperate and no player 2s attend to looking.

Let β_{CWL} represent the proportion of player 1s who cooperate with looking. We classify as CWL any population state in which $\beta_{CWL} + \beta_{CWOL} = 1$ and $\delta_{EID} = 1$.

ALLD: We categorize as ALLD all the states where no player 1s cooperate and most player 2s end unconditionally.

Let β_D represent the proportion of player 1s who always defect and δ_D be the proportion of player 2s who always end. We classify as ALLD any population state in which $\beta_D = 1$,

$$\begin{aligned} pc_l + (1-p)c_h &\geq (1-\delta_D) \cdot \frac{a}{1-w} + \delta_D \cdot a, \\ pc_l + (1-p)c_h &\geq \delta_{EID} \cdot \frac{a}{1-w} + (1-\delta_{EID}) \cdot a, \text{ and} \\ c_l + (1-p)c_h &\geq \delta_{EIL} \cdot (ap + c_h(1-p)) + \delta_{EID} \cdot \left(ap + c_h \frac{1-p}{1-pw} \right) + \delta_D \cdot (ap + c_h(1-p)) \, . \end{aligned}$$

These conditions' interpretation is that the payoffs to player 1 from always defecting are greater than the payoffs from CWOL, CWL, or ONLYL given the proportion of player 2's who are ending if player 1 looks or ending if player 1 defects. These condition ensure that player 2s are ending sufficiently often that player 1s do not benefit from deviating to other strategies.

C.2 Figure 3 Simulation Details

For the simulations presented in Fig. 3 of the manuscript, we employ the following parameters: $p = 0.51, b = 1.0, c_l = 4.0, c_h = 12.0, d = -10.0, w = 0.895$. We let a vary from 0 to 2, in intervals of 0.04. For each value of a, we perform 10,000 simulations. In each simulation, we randomly seed the strategy frequencies, adjust strategy frequencies according to the replicator equation for 1,000 time periods–enough time for the population to stabilize–then record the frequency of each strategy. Since the replicator equation is an ODE that cannot be solved analytically, throughout, we numerically estimate the replicator equation using Matlab's ode45 function.

D The Role of the Assumption $pb + (1-p)d < 0$

In the manuscript, we claim that, "the inequality $pb+(1-p)d < 0$ must hold for CWOL to emerge." We support this claim in two ways. First, we relax this assumption and identify a fourth equilibrium in the region where $-d \leq \frac{p}{1-p}b$. It is the strategy pair where player 1 looks and cooperates whenever the temptation is low as long as she has always done so in the past and defects otherwise, and player 2 ends iff player 1 has ever defected when the temptation is low (we refer to this as the ONLYL equilibrium). We now prove that this equilibrium is subgame perfect. Notice that for the following claim we must assume that player 2 is informed of the temptation level at least when player 1 is informed.

Claim 13. *Iff $-d \leq \frac{p}{1-p}b$ and $a + (pa + (1-p)c_h) \cdot w/(1-w) \geq c_l$, there is a subgame perfect equilibrium where player 1 looks and cooperates whenever the temptation is low as long as she has always done so in the past, and player 2 continues iff player 1 has always cooperated when the temptation is low.*

Proof. We begin by checking player 1's potential deviations. Player 1's first possible deviation is to cooperate if the temptation is high. The payoff from this deviation is $(a - c_h) \cdot 1/(1-w)$, which is negative by assumption. Player 1's second possible deviation is to defect if the temptation is low. The payoff from this deviation is c_l, whereas the payoff from maintaining the strategy specified by the equilibrium and cooperating is $a + (pa + (1-p)c_h) \cdot w/(1-w)$. Thus, player 1 will not deviate iff $a + (pa + (1-p)c_h) \cdot w/(1-w) \geq c_l$. Now suppose player 1 finds herself in the history where she has deviated in the past and player 2 has nevertheless continued. Player 1 prefers to defect in the low temptation state, since player 2 will end regardless, and cooperation is costly by assumption.

Next, we consider player 2's potential deviations. Player 2's only possible deviation is to end. If she does so, she gets 0, whereas if she continues as prescribed by the equilibrium, she gets $1/(1-w) \cdot [pb + (1-p)d]$. Player 2 prefers to continue as prescribed by the equilibrium iff $-d \leq \frac{p}{1-p}b$. Finally, suppose player 2 finds herself in the history where player 1 has deviated. If player 2 ends as prescribed, she gets 0 in the future, while if she continues, player 1 will defect when the temptation is low, and player 2's expected payoff is $\frac{d}{1-w} < 0$, so she does not benefit from deviating. □

Second, we perform replicator simulations over a range of values of d, and expand the set of strategies available to player 2 so that it includes "end if player 1 defects when the temptation is low." Payoffs in this game are represented in Fig. S2

We classify population frequencies as behaviorally equivalent to the ONLYL equilibrium if nearly all (> 0.999) player 1s cooperate only when the temptation is low, and nearly all (0.999) player 2s end if player 1 defects when the temptation is low. Note that we redefine the equilibrium classification accordingly. The results of the simulations are presented in Fig.S3. ONLYL emerges often in the region where it is an equilibrium, whereas CWOL rarely emerges. When $-d > \frac{p}{1-p}b$, ONLYL does not emerge; the only cooperative equilibrium that emerges is CWOL.

E Cooperating Without Looking at the Probability of Continuation

In the Envelope Game, the cost of cooperation varies, and player 1 learns this cost when she looks. Some behaviors are better explained by a model in which another parameter varies, and player 1 learns this other parameter when she looks. One-shot giving is one such behavior; as we suggest in the manuscript, one-shot giving can be explained by a model in which the probability of continuation varies and player 1 learns this probability when she looks. We demonstrate this now with the following variation of the Envelope Game.

In each round, player 1 receives an envelope, which contains the continuation probability w. w is 1 with probability p and 0 with probability $1-p$. Player 1 can choose to open the envelope and thus find out w, or not. Then player 1 decides to cooperate or to defect. Subsequently player 2 can either continue or end the game. In the former case, there is another round with probability w. If player 1 cooperates, her payoff is a, while player 2 receives b. If player 1 defects, her payoff is c, which does not vary, while player 2 receives d. We assume $c > a > 0$, and also retain the assumptions $b > 0 > d$ and $pb + (1-p)d < 0$.

In this game, cooperating without looking is an equilibrium in the parameter region $\frac{a}{1-p} > c$. The left hand side of this equation is the expected payoff of a long-term cooperative interaction which, on average, continues with probability $p \cdot 1 + (1-p) \cdot 0 = p$. However, cooperating with looking is never an equilibrium. This is because if player 1 looks, in periods when $\delta = 0$, she cannot benefit from cooperating. Notice that player 1s who cooperate without looking will sometimes cooperate in one-shot situations, and this may explain why we observe one-shot giving, i.e. when $w = 0$.

References

[1] Aumann, R. J. Correlated equilibrium as an expression of bayesian rationality. *Econometrica: Journal of the Econometric Society* 1–18 (1987).

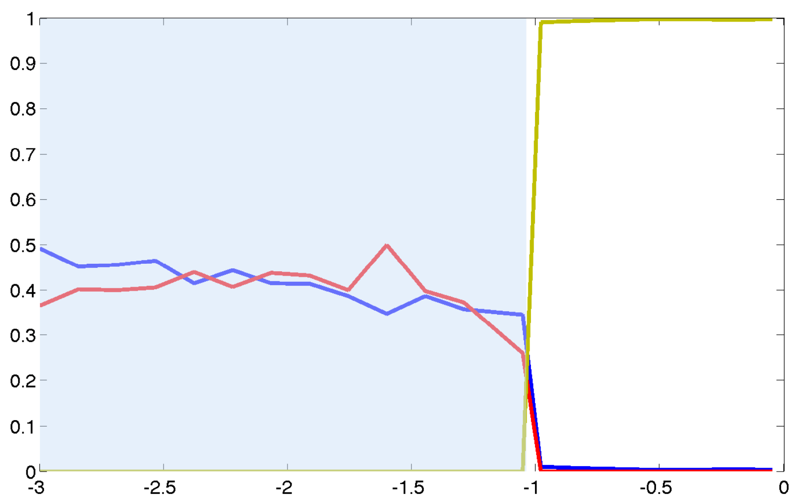
[2] Osborne, M. J. & Rubinstein, A. *A course in game theory* (MIT press, 1994).

[3] Aumann, R. Survey of repeated games. In Bohm, V. (ed.) *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, 11–42 (Bibliographisches Institut, Mannheim, 1981).

Figure S2: Payoffs in the Restricted Envelope Game when player 2 can end if player 1 defects when the temptation is low

Player 1	Player 2: End if Player 1 Looks	End if Player 1 Defects	Always End	End if Player 1 Defects When the Temptation is Low
Cooperate Without Looking	$\frac{a}{1-w}, \frac{b}{1-w}$	$\frac{a}{1-w}, \frac{b}{1-w}$	a, b	$\frac{a}{1-w}, \frac{b}{1-w}$
Cooperate With Looking	a, b	$\frac{a}{1-w}, \frac{b}{1-w}$	$\frac{a}{1-w}, \frac{b}{1-w}$	
Look and Cooperate Only When Temptation is Low	$ap + c_h(1-p), bp + d(1-p)$	$\frac{ap+c_h(1-p)}{1-pw}, \frac{bp+d(1-p)}{1-pw}$	$ap + c_h(1-p), bp + d(1-p)$	$\frac{pa+(1-p)c_h}{1-w}, \frac{pb+(1-p)d}{1-w}$
Always Defect	$c_l p + c_h(1-p), d$	$c_l p + c_h(1-p), d$	$c_l p + c_h(1-p), d$	$\frac{c_l p+c_h(1-p)}{1-(1-p)w}, \frac{bp+(1-p)d}{1-(1-p)w}$

Figure S3: Learning dynamics of the Envelope Game when the harm to player 2 from defection is varied



We run 500 time series with randomly seeded strategy frequencies for a range of values of d, and record the frequency with which they stabilize in outcomes which correspond with the Nash equilibria of the envelope game, as discussed in section C.1. We vary the value of d along the x-axis. The y-axis represents frequencies, and each colored line presents the frequency of the strategy pair: ALLD is represented in red, CWOL in blue, CWL in green, and ONLYL in yellow. The parameter region in which CWOL is the only cooperative equilibrium is shaded in light blue.