Title: Local norms of cheating and the cultural evolution of crime and punishment

Authors: K.B. Schroeder[1,2,3], G. Pepper[2], D. Nettle[2]

Corresponding author: K.B. Schroeder, +1 (650) 387-3993, kari.britt.schroeder@gmail.com, Department of Anthropology, University of California, Davis, One Shields Ave, Davis, CA 95616, USA

[1]Department of Anthropology, University of California, Davis

[2]Centre for Behaviour and Evolution, Newcastle University

[3]Department of Psychology, Boston University

Summary.

While evolutionary theorists have studied how prosocial behaviors can spread, social scientists have examined the correlates of neighborhood crime. To integrate these endeavors, we conducted an experimental game of theft in two adjacent neighborhoods, one low in crime (A) and the other crime-ridden (B). We investigated how decisions to engage in or sanction antisocial behavior are shaped by, and shape, the local sociocultural environment. Participants reported their perceived neighborhood frequency of cheating on public goods. They played a third-party punishment (3PP) game used to assess theft, 3PP of theft, and expectation of 3PP. Participants in B thought their neighbors commonly cheat but did not condone cheating. They stole more money, and were less punitive of theft, than those in A. Perceived cheating was associated with theft (positively) and expectation of 3PP (negatively) and central to the neighborhood difference. We examined the causality of perceived cheating with a norms manipulation. Residents in B who were informed that cheating is locally uncommon were more expectant of 3PP. The perception that cheating is common can lead to further antisocial behavior via three feedback mechanisms. Consideration of these and of norm psychology will help us understand how neighborhoods get stuck in an antisocial culture.

Key words.

cooperation, social disorganization theory, social capital, descriptive norms, injunctive norms

Introduction.

Why do humans contribute to the public good? Proximately, one explanation is punishment of free-riding [1]. Empirical evidence for this comes from economic games. Using a repeated public goods game, Fehr and Gächter showed that the opportunity for players to fine each other on the basis of contribution behavior can stabilize contributions to the public good at a high level [2]. Following this, the large-scale covariation of cooperative behavior and punishment targeted at non-cooperative behavior has received substantial interest [3][4]. Considerable local variation in cooperative behavior has also been observed [5–7]. Whether harsher punishment covaries with more cooperative behavior at the local level as well has as yet spurred little research among students of the evolution of human behavior and experimental economics (but see [8]).

However, this question has generated substantial research within the fields of sociology and criminology, which are concerned with local variation in provisioning of a public good: a crime-free neighborhood. Social disorganization theory posits that inequality (e.g., poverty, residential mobility, family disruption) can result in communities characterized by reduced participation in local organizations and fewer local friendships. This low 'social capital' can lead to increased crime via reduced efficacy for collective action; i.e., residents lack the capacity to enforce desirable behavior and maintain a crime-free neighborhood [9–11][12]. Sampson et al. emphasize the necessity of trust and shared behavioral expectations for informal social control (i.e., informal surveillance and/or intervention by residents) [13]. Support for the association between low social capital and high crime rates comes primarily from survey and administrative data [13]. However, whether social control is the process that unites these two phenomena is largely speculative, as data on actual social control (rather

than the potential for social control) are difficult to come by [12,14].

In this paper, we present two studies that integrate ideas from the behavioral economic, evolutionary, sociological, and social psychological literatures. Integration of these approaches assists us in investigating correlations between individual decision-making, as assessed with experimental economics, and the local sociocultural environment. Our results suggest three reciprocal processes that link antisocial behavior at the individual and neighborhood level, potentially leading to a downward spiral of crime and disorder.

Our studies were set in two neighborhoods in Newcastle Upon Tyne, England, which differ dramatically in crime rates and socioeconomic deprivation. While Neighborhood A is relatively affluent, nearby Neighborhood B has experienced high rates of unemployment, physical decay, massive depopulation, and crime, following the collapse of mining and shipbuilding industries (see [6] and citations therein). In an earlier study, we used surveys, a Dictator Game, observation of antisocial behavior, and field experiments to reveal substantially less antisocial behavior and more social capital and prosocial behavior in Neighborhood A than B [6]. Here, we return to these neighborhoods to investigate how social capital and local cooperative norms relate to crime-like behavior and punishment in an economic game. *Study 1* was observational and aimed to document and explain differences in behavior between the neighborhoods. *Study 2* introduced a novel experimental methodology to manipulate perceived norm adherence and study the consequences of doing so.

Study 1.

Camerer and Fehr suggest that a real-world example of a third-party punishment game (3PP game) [15] is scolding of a neighbor for treating another person unacceptably [16]. In this

study, we administered a 3PP game along with a questionnaire (see Electronic Supplementary Material (ESM)). Our variant of the game, which was played among residents within each neighborhood, enabled us to study differences between the neighborhoods in stealing and punishment for stealing. Player 1 was given the opportunity to steal from Player 2. Player 3 was given the opportunity to fine Player 1 if she took money from Player 2. Player 2 indicated whether she thought Player 3 would fine Player 1 if Player 1 took half of her money.

We used Player 3 game behavior and the questionnaire to assess 1) whether residents of Neighborhood A were more willing than those of B to punish antisocial behavior in their neighborhood, and 2) whether neighborhood trust can explain the hypothesized relationship between neighborhood and punishment behavior.

If punishment for antisocial behavior is effective in maintaining a crime-free neighborhood, then residents must be adjusting their behavior with respect to cues that antisocial behavior will not be tolerated. What are the salient cues that antisocial behavior will not be punished? We consider the role of social norms in behavior. Cialdini et al. distinguish between *injunctive* and *descriptive* norms [17]. Injunctive norms convey how people should behave, while descriptive norms illustrate how most people do behave. It is the ability of the neighborhood to induce its members to adhere to injunctive norms, i.e. to “regulate its members according to desired principles” [11], that proponents of social disorganization theory consider threatened in the presence of low social capital. A lack of alignment between descriptive and injunctive norms is also implicit in the idea that signs of social and physical disorder invite criminal behavior [18][19]. This mechanism for the 'spread of disorder' was elegantly tested by Keizer et al. [20], who created public spaces in which an

explicit injunctive norm was violated – e.g., a littered space next to a sign telling people not to litter – thereby communicating a lack of adherence to the injunctive norm. This treatment induced further antisocial behavior that was not specific to the violated norm. A largely untested interpretation of these results is that signs that others are flouting injunctive norms may serve as cues that antisocial behavior will not be punished (but see [21], who investigated an association between neighborhood physical disorder and expectation of arrest for a crime).

Thus, we used Player 1 and Player 2 behaviors to assess whether residents of Neighborhood B were 1) more likely to behave antisocially and 2) less likely to expect someone in their neighborhood to intervene in antisocial behavior. We also asked about injunctive and descriptive civic norms to determine whether a perceived lack of neighborhood civic norm adherence could explain the hypothesized difference between neighborhoods in 3) antisocial behavior and 4) expectation of 3PP for antisocial behavior.

Study 1 Methods.

Sampling. The Ethics Committee of the Newcastle University Faculty of Medical Sciences approved the study protocol. We conducted the study from July 2012 to December 2012. A maximum of one participant per household was drawn from the electoral roll. Potential participants received a hand-delivered envelope with a cover letter describing the study, packet (questionnaire, explanation of the game, and game), and stamped return envelope. We avoided sampling adjacent households and households sampled by [6].

Questionnaire. From the questionnaire, we recorded each participant's *age* and sex (*male*).

Trust. We asked individuals how much they trust people in their neighborhood, on a 10-point scale (10 = most trusting).

Civic norms: condoned and perceived cheating. We asked individuals about both injunctive and descriptive civic norms (ESM). For the injunctive norm, we described three behaviors and asked whether it is Never OK to do this behavior, Always OK, or somewhere in between. Answers were constrained to a 10-point scale (1 = 'Always OK' and 10 = 'Never OK'). The behaviors were 1) cheating the benefits system, 2) avoiding a fare on public transport, and 3) cheating on taxes. *Condoned cheating* is the average across behaviors. Larger values indicate that cheating on public goods is condoned. Note that *condoned cheating* is similar to the 'norms of civic cooperation' [22][3] derived from the World Values Survey.

For the descriptive norm, we asked individuals whether they think many people in their neighborhood would do this behavior (1 = 'No one would' and 10 = 'Everyone would'). We averaged across these responses to arrive at *perceived cheating*. Larger values indicate that neighborhood cheating on public goods is perceived as more common.

The 3PP game. Participants read instructions for the game, which followed the questionnaire, and then worked through examples (see ESM). (From this, we had responses to six *test questions*.) They were told that after receiving the packet in the post, we would determine the game outcome and then deliver their cash payoff along with a £5 payment for completing the survey.

The game worked as follows: All three players received a hypothetical £10. Player 1 had to decide how many pounds (integer from 0 to 10) to take from Player 2. If Player 1 took

money from Player 2, Player 3 had to decide whether to fine Player 1. We used the strategy method for Player 3. Player 3 had to decide, for each amount greater than 0 that Player 1 could take, whether to pay to fine Player 1. Therefore, Player 3 had to make 10 choices, each corresponding to an amount that Player 1 might take from Player 2. The cost of the fine to Players 1 and 3 was constant (Player 3 paid £2 to make Player 1 lose £6). Player 2 could not make a choice in the game. We asked Player 2 to indicate whether she thought Player 3 would fine Player 1 if she took £5 from her (ESM).

Game behaviors are thus: *theft* (an integer from 0 to 10 representing the amount of money Player 1 took from 2), *expect 3PP* (whether Player 2 expected Player 3 to punish Player 1 if she took £5), and *punitiveness* (an integer from 0 to 10; this is the total number of potential thefts, from one pound £1 to £10, that Player 3 would punish).

Subjective value of money. We expected the subjective value of money to differ between neighborhoods and impact game behavior. Therefore, following the game, we asked how much of a difference, on a scale of one to 10, an amount of money *x* would make to their weekly budget, where *x* was £1 for Player 1 (*difference £1 makes*) and £2 for Players 2 and 3 (*difference £2 makes*). After commencement of data collection, we revised the packets for Player 1 to include *x* = £10. Thus, for some Player 1s we also have *difference £10 makes*.

Statistical analyses. The majority of responses can be considered discrete ordered choices. Thus, to assess neighborhood differences in game behavior, trust, civic norms, and the value of money, we analyzed the data with ordered logistic regression. The exception to this is game behavior for Player 2, for which we used binary logistic regression. We compared the

fit of different models with the Akaike information criterion (AIC) [23]. Ordered and binomial logistic regression analyses and plotted predictions were produced in the R statistical and computing environment [24] with the following packages: MASS [25], rethinking [26], beeswarm [27], and ggplot2 [28]. Note that plotted predictions for *theft* and *punitiveness* are both (0, 8). For each of these game behaviors, two possible values were not observed (3 and 8 for *theft*, 2 and 9 for *punitiveness)*; thus, for prediction we condensed the ranges. We report Odds Ratios (ORs) for a unit increase in the outcome for each unit increase of the predictor variable, accompanied by 95% confidence intervals.

Study 1 Results.

Participants. We achieved sample sizes of 40 (16 male), 44 (22 male), and 49 (23 male) for Players 1, 2, and 3, respectively, in Neighborhood A and 34 (12 male), 43 (23 male), and 50 (23 male) in B (Table S1). Every week, new players from each neighborhood were combined into triads, and we determined game outcome from their decisions. For incomplete triads, players were drawn at random from all previous neighborhood players. We delivered to participants: the game outcome, debriefing sheet, money received from the game, and £5 for participating. The mean payoff from the game is £9.26 (σ = £3.49) in Neighborhood A and £9.16 in B (σ = £4.13). Descriptive statistics and neighborhood comparisons for key variables are in Table S2.

Trust. Participants in Neighborhood A indicated far higher *trust neighbors* than did participants in B (Table S2) (OR 18.8, 95% CI 10.8-32.8).

Punishment of antisocial behavior. As predicted, participants in Neighborhood A were more punitive than those in B (Figure S1) (OR 3.3, 95% CI 1.6-7.0). Median *punitiveness* is 6 (MAD = 4) and 3 (MAD = 3) for Neighborhoods A and B, respectively. Thus, more participants in Neighborhood A indicated that they would pay £2 to fine Player 1 for a greater number of potential thefts.

The subjective cost of punishment in the game, *difference £2 makes*, had a negative effect upon *punitiveness* (OR 0.7, 95% CI 0.6-0.9) and was larger for participants in Neighborhood B than A (Table S2). However, participants in Neighborhood A were still more punitive than those in B when we include *difference £2 makes* in the model (OR 2.1, 95% CI 0.9-4.6). This result is robust to the inclusion of additional covariates *age*, *male*, and *test questions* (OR 2.9, 95% CI 1.2-7.2).

We hypothesized that greater trust among residents of Neighborhood A would partially explain the increased willingness of residents to engage in 3PP of antisocial behavior. Individuals who reported greater *trust neighbors* were slightly more punitive (OR 1.15, 95% CI 0.99-1.32). The relationship between trust and punitiveness is not robust to the inclusion of *difference £2 makes*; however, including an interaction between *difference £2 makes* and *trust neighbors* improves model fit (AIC of 380.13 compared to 384.49).

Predictions from the model including the interaction are shown in Figure 1. *Difference £2 makes* still has a negative effect on *punitiveness*, but the slope is steeper for participants with high *trust neighbors*. Thus, participants with high *trust neighbors* are more punitive than those with low *trust neighbors* when *difference £2 makes* is small, but less punitive when it is large. *Neighborhood* is no longer a reliable predictor of *punitiveness* when the interaction is included in the model (OR 1.8, 95% CI 0.7, 5.7), nor does model fit

improve with the addition of *neighborhood* (AIC = 380.67).

Civic norms: condoned and perceived cheating. We observed little variation in assessments of injunctive norms across civic behaviors (Table S2). Nor did we detect a clear difference between neighborhoods with respect to either the within-participant mean of injunctive norms, *condoned cheating*, or specific injunctive norms (Figure 2, Table S2). Thus, in both neighborhoods, most participants indicated that it is not acceptable to cheat on public goods.

However, there was a dramatic difference between neighborhoods with respect to *perceived cheating*, as well as each of the specific descriptive norms. More participants in Neighborhood B indicated that more of their neighbors would cheat on a public good than those in A (Figure 2, Table S2). Participants who thought more of their neighbors cheat on public goods were also less trusting of their neighbors (OR 0.54, 95% CI 0.48-0.62).

Juxtaposition of *condoned cheating* and *perceived cheating* reveals that although participants in Neighborhood B tended to state that many of their neighbors cheat on public goods, we lack strong evidence that they view this behavior as more acceptable than those in A. We therefore use *perceived cheating* as a within-participant measure of the perceived lack of neighborhood civic norm adherence.

Antisocial behavior.

Participants in Neighborhood B took more from their neighbors in the game. *Theft* is also more variable in Neighborhood B than A. The median value of *theft* is 5 in Neighborhood B (MAD = 5), compared to 0 in A (MAD = 0) (odds that *theft* is greater in Neighborhood B: OR 2.9, 95% CI 1.2-7.1). The neighborhood difference in *theft* is robust to the inclusion of *age*,

male, and *difference £1 makes* (OR 2.8, 95% 2.5-6.9). For the reduced dataset for which we had data on *difference £10 makes* (40 participants, 23 from Neighborhood A), substituting this variable in the model increases the odds that a participant in B stole more in the game (OR 4.1, 95% CI 0.9-17.5). Inclusion of *test questions* in the model reduces confidence in the neighborhood difference in *theft* (OR 2.1, 95% 0.8-5.8). However, incomplete test questions are heavily patterned for Player 1; only participants in Neighborhood B for whom *theft* > 0 did not complete the questions. Irrespective of the participant's comprehension of the entire game, the opportunity for Player 1 to behave antisocially (the outcome of interest to us) should be very clear from the packet (i.e., "How many pounds do you choose to take from Player B? ") (ESM).

As expected, *perceived cheating* is a robust predictor of *theft*, even controlling for *difference £1 makes* (Figure 3; OR 1.3, 95% CI 1.0-1.6). When both *neighborhood* and *perceived cheating* are considered in the same model, neither is a reliable predictor of *theft*. Nor does AIC offer strong support for a single model (235.40 for the model with *perceived cheating*, 234.67 for *neighborhood*, and 234.60 for *perceived cheating* + *neighborhood*). This suggests that to understand the greater *theft* in Neighborhood B, we need to consider *perceived cheating*.

Expectation of 3PP. We asked Player 2 whether she thought Player 3 would fine Player 1 if Player 1 took £5 from her (*expect 3PP*). Contrary to our expectations, *neighborhood* was not a reliable predictor of *expect 3PP*. Of participants in Neighborhood A, 36.36% expected 3PP, compared to 30.23% of participants from Neighborhood B (OR 1.2, 95% CI 0.5-3.2). However, as predicted, we did observe a negative relationship between *perceived cheating*

and *expect 3PP* (Figure 4; OR 0.8, 95% CI 0.6-1). This relationship does not change with inclusion of *difference £2 makes* as a proxy for the local subjective value of £2 (OR 0.8, 95% CI 0.6-1).

Study 1 Summary and Discussion.

Participants in Neighborhood B were far more likely than those in A to think that more of their neighbors cheat on public goods. We could not, however, attribute this to residents of Neighborhood B condoning cheating more than those in A. Thus, a perceived lack of civic norm adherence was pervasive in Neighborhood B. Correspondingly, participants in Neighborhood B indicated far less trust in their neighbors than did those in A. This result fits with the far lower self-reported social capital in Neighborhood B we previously observed (our measure of trust in the current study, *trust neighbors*, approximates one of six items in our social capital index [6], which was highly positively correlated with the overall index (0.77, p-value < 0.05) [6]).

The lower trust in Neighborhood B would appear justifiable: participants in B stole more money and were less punitive of theft in the game. However, our results suggest that the greater theft in the game in Neighborhood B may be partly due to the common perception that cheating is prevalent. Our observation that perceived neighborhood cheating is positively associated with stealing in the game is in accordance with the results of [29], who demonstrated a positive effect of observed theft on a participant's subsequent choice to steal in the lab, as well as those of [17,20], who showed that observed norm violation can result in an increase in norm violation.

As expected, *trust neighbors*, a core component of social capital, was a positive

predictor of punitiveness. Kocher et al. similarly found that trust in members of a participant pool was positively correlated with punitiveness in a public goods game [8]. Although they interpreted this outcome as stemming from greater disappointment in free-riding behavior, they suggest it merits further investigation of the role of social capital in norm enforcement.

One interpretation of the unexpected interaction we observed between *trust neighbors* and *difference £2 makes* lies in consideration of the multiple ways in which the cost of punishing can vary for the punisher. We showed that participants were more punitive when *difference £2 makes* was smaller. Punitiveness is also less costly when there are fewer defectors and/or more punishers [30–32]. *Trust neighbors* – which was associated with greater punitiveness - may be informative as to whether Player 3 thinks there are many punishers and defectors in her neighborhood and thus construed as a measurement of the cost of intervening in antisocial behavior. From this perspective, our results are consistent with the idea that people are more punitive when punishment is cheap – with respect to both material resources and the behavior of others.

We are unable to determine whether participants in Neighborhood B stole more money than those in A because they thought punishment was less likely. This is because a participant's motivation to steal a particular amount of money can be ascribed to inequity aversion as well as the expected probability of punishment. However, our data from Player 2 addresses expectation of punishment. While we did not observe a robust neighborhood difference in *expect 3PP*, we did observe a strong negative relationship between *perceived cheating* and *expect 3PP*. That is, a participant who thought many of her neighbors cheat on public goods was less likely to expect a neighbor to pay £2 to fine Player 1 if she took half her money.

This result supports the idea that expectations of social punishment are conditioned on the believed frequency of norm violation [33]. It also suggests that expectation of punishment is one of the mechanisms by which signs of norm violation can lead to further violation [18,33]. However, the causality of the relationship between *perceived cheating* and *expect 3PP* remains uncertain. Surveys of the kind in *Study 1* can only establish correlation; examining the causal significance of one variable for another requires experimental manipulation of the first variable. With this in mind, we undertook *Study 2*, in which we used selective feedback from *Study 1* to experimentally alter perceptions of local norm adherence in the two neighborhoods.

Study 2.

Feedback on or manipulation of descriptive norms has been used to alter people's behavior – from littering [17] to energy use [34]. In *Study 2*, we used a novel method for manipulation of descriptive norms to investigate the causality of the relationship between *perceived cheating* and *expect 3PP*. In each neighborhood, we provided novice Player 2s with information on what their neighbors thought about the descriptive norms of the neighborhood ('Norms treatment'). We manipulated this information so as to present *Study 2* participants from Neighborhood A with a less positive picture of perceived descriptive norms than was really the case, and participants from Neighborhood B with a more positive picture. We predicted that participants in Neighborhood A who received the Norms treatment would be less likely to expect Player 3 to 3PP on their behalf, compared to those participants in the same neighborhood who did not receive the treatment. We predicted the opposite effect in Neighborhood B.

Study 2 Methods.

Sampling. We collected data for *Study 2* from October to December 2012, while *Study 1* was ongoing (ESM), following the same protocol as in *Study 1*.

Norms Questionnaire. We refer to the questionnaire used in *Study 1* as 'Baseline treatment'. The questionnaire for the Norms treatment differed as follows.

Civic norms manipulation: perceived cheating. The Norms questionnaire did not include questions about injunctive and descriptive norms. We presented participants with information on the responses of a subset of *Study 1* participants in their neighborhood to the questions about descriptive civic norms (ESM).

The following backstory was used: As a part of the Tyneside Neighbourhoods Project, we had asked 10 people in their neighborhood how common they think avoiding a public transport fare, cheating the benefits system, and cheating on taxes, are in that neighborhood. We averaged these answers to get an idea of how common people think certain behaviors are. We wanted to know what other people in the neighborhood thought of these answers, and thus were asking them (ESM).

We presented one scale for each of the behaviors. The information in each scale was manipulated: in Neighborhood A, we took the mean of the 10 responses that gave the least favorable impression of cheating (i.e., high *perceived cheating*), and in Neighborhood B, we took the mean of the 10 responses that gave the most favorable impression of cheating (i.e., low *perceived cheating*). The information presented for Neighborhood A was: 5.7 for avoid a fare on public transport, 5.5 for cheat the benefits system, and 6.7 for cheat taxes (1 ='No one would', 10 = 'Everyone would'). In Neighborhood B it was: 2.2 for avoid a fare, 2.3 for cheat

benefits, and 1.7 for cheat taxes. Beneath each scale, *Study 2* participants were asked to circle 'Fewer people would do this,' 'This is about right,' or 'More people would do this' (ESM).

Contamination. To assess whether participants knew *Study 1* participants, we included a contamination question: 'Do you know of other people in your neighborhood who got a questionnaire and plan to post it or already have posted it?' ('Yes', 'Not sure', or 'No').

3PP game. For *Study 2*, we measured the following behavior: *expect 3PP* (yes or no; representing whether Player 2 expected Player 3 to punish Player 1 if she took £5 from her).

Statistical analyses. We used binary logistic regression to assess the effect of the Norms treatment on *expect 3PP* within each neighborhood.

Study 2 Results.

Participants. For *Study* 2, we sampled 41 participants from Neighborhood A (21 male) and 39 participants from B (16 male) (Table S3; ESM).

Reaction to normative information. Participants in Neighborhood B were far more likely than those in A to indicate 'This is about right' when presented with the manipulated norms scales for *cheat benefits* and *avoid fare* (OR 3.63, 95% CI 1.23-10.70 and OR 3.74, 95% CI 1.34-10.49, respectively). In Neighborhood B, 38.46%, 43.59%, and 46.15% of participants indicated 'This is about right' for *cheat benefits*, *avoid fare*, and *cheat taxes*, respectively. The majority of participants in Neighborhood A indicated 'Fewer people would do this' when presented with the manipulated scales for *cheat benefits* and *avoid fare* (78.05% of

participants for each behavior). Only 51.28% of participants in Neighborhood A indicated 'Fewer people would do this' for *cheat taxes.*

Expectation of 3PP: Norms treatment. Participants in Neighborhood B who received the Norms treatment – i.e., who received information that their neighbors perceive cheating to be uncommon – were more likely to expect Player 3 to 3PP on their behalf, compared to those in B who received the Baseline treatment. The proportion of participants who expected 3PP is 58.97% for the Norms treatment, compared to 30.23% for Baseline (OR 3.32, 95% CI 1.33-8.25; Figure S2). Exclusion of participants for whom *contamination* was 'Not sure' (five) or 'Yes' (two) does not qualitatively change the results. (One participants circled both.)

We did not observe a robust effect of the Norms treatment on *expect 3PP* in Neighborhood A. Contrary to our prediction, the proportion of participants in A who expected 3PP is 41.46% for Norms treatment, compared to 36.36% for Baseline treatment (OR 1.24, 95% CI 0.52-3.00; Figure S2).

However, the Norms treatment generated an unanticipated response in Neighborhood A. Some participants attempted to redirect their money by asking us to: donate it money to charity (three participants), keep it for research/university funds (two), or not pay them (one). The rate of 'opting out of payment' is 14.63% for Norms treatment participants in Neighborhood A, compared to 1.15% of Baseline participants in A (OR 11.25, 95% CI 2.18-57.97). This spontaneous change in game play was never observed in Neighborhood B.

Study 2 Summary and Discussion.

In *Study 2*, participants in Neighborhood B received information that their neighbors think

there is little cheating on public goods in their neighborhood, relative to what we actually observed in *Study 1*. They were far more likely to expect a neighbor to 3PP antisocial behavior compared to those in B who did not receive the manipulation. Whether disorder can play a causal role in an increase in crime rates [18] has been debated [13,35]. Our results provide empirical evidence of a mechanism by which norm violation can lead to the further violation of a different norm – through change in the expectation of punishment.

We did not observe a reliable negative effect of the descriptive norms manipulation on expectation of 3PP in Neighborhood A. It is not clear why we observed the expected result in Neighborhood B and not A. In *Study 1*, we found greater variation in trust and norms in Neighborhood B than in A (Table S2). One interpretation of this is that the environment is more heterogeneous and unpredictable in Neighborhood B. If so, perhaps residents of Neighborhood B are less certain than residents of A of the behavior of their neighbors and therefore were more accepting of the manipulation. Indeed, far more Neighborhood B participants circled 'This is about right' when presented with the manipulated descriptive norms. Another possibility is that participants in Neighborhood B were more accepting of the information provided by an authority figure (university personnel/scientist).

General Discussion.

In *Study 1*, we demonstrated that the covariation of cooperation and punishment of non-cooperation, which has been observed cross-culturally with economic games [4], can extend to the local level. Participants in Neighborhood A stole less money and were more punitive in the game than those in B. The lower theft in the game in Neighborhood A corresponds to our previous observation that residents of A were drastically more cooperative by most measures

than those of B [6]. Thus, Neighborhood B appears to be characterized by relatively low cooperation, low trust, and low willingness to intervene when others are behaving antisocially. In *Study 2*, we showed that providing participants in Neighborhood B with information that cheating is perceived as uncommon within their neighborhood led to a sharp increase in the expectation of third-party punishment for theft-like behavior. An increase in the perceived likelihood of punishment would presumably lead to greater cooperation, given the close relationship between these two variables. Thus, our results provide novel empirical support for a mechanism by which cues of norm violation can lead to further norm violation [17,20]: altered expectation of punishment [18,33].

We consider these results within a framework where culture is dynamic, subject to evolutionary processes that can lead to more or less cooperative outcomes [36]. The apparent disparity between desired and achieved cooperative outcomes in Neighborhood B, as assessed by the discrepancy between condoned and perceived cheating, adds new perspectives on the cultural evolution of variable cooperative outcomes. Unlike in recent cross-cultural studies of cooperation and punishment [3,4], our two study populations share many cultural components, including the institutions that formally sanction their civic violations (although *how* those institutions are experienced may vary) and injunctive civic norms. Thus, we are in a unique position to ask precisely what the salient differences are that may limit Neighborhood B achieving a higher level of cooperation.

One of the most striking differences between Neighborhoods A and B is in perceived frequency of cheating on public goods. *Perceived cheating* (to the extent that it reflects real differences in neighborhood frequencies of cheating) should not be considered as solely an outcome variable that shares a common origin with theft in the game or crime and disorder in

real life (cf. [37]). The correlational analyses of *Study 1* and the experimental manipulation of *Study 2* suggest that perceived cheating on public goods stands in a causal relationship to game behaviors.

In particular, our results reveal three potential routes by which perceived cooperative norm violation can lead to further violation of cooperative norms. 1) To avoid being 'suckered', conditional cooperators are motivated to defect if they perceive that defection is common [29,38–40]. 2) Perceived cheating leads to lower trust - and low trust leads to reduced informal punishment of norm violation. Similarly, extensions of social disorganization theory include feedback processes between crime/disorder and social cohesion/control, via fear or residential instability [14,35,37]. Traxler and Winter also observe a direct effect of the perceived frequency of norm violations on expressed willingness to sanction violations [33]. 3) When the perceived frequency of cooperative norm violation is high, expectation of punishment for violation is lower [41]. All of these mechanisms may be partial explanations for the interdependence of individual decisions to engage in crime [42].

We hypothesize that these three positive feedback mechanisms, wherein perceived cooperative norm violation leads to further cooperative norm violation, could act simultaneously to result in a rapid downward spiral, leading to low levels of cooperation. As Cialdini et al. note [17], descriptive norms are informative as to adaptive behavior. In a community with low levels of cooperation and minimal sanctioning for cooperative norm violation, non-cooperative strategies may outperform others [43]. Other processes - prestige-biased [44] or conformist [45] transmission and self-selection of people with preferences for an antisocial culture the community - could further reinforce uncooperative strategies. While

cooperative norms are considered a component of social capital [22,46], our results demonstrate the need for explicit integration of cultural transmission and norm psychology (that is, psychological adaptations for determining and adopting local norms and punishing violators [47]) with social disorganization theory. Scholars of criminology will note the similarities between the former and the social learning theory of deviance [48]. However, we extend this bridge between the sociocultural environment and individual behavior by emphasizing the feedback from the individual to the sociocultural group. That is, we have outlined three routes by which an individual's defection can lead other individuals to adopt similar behavioral strategies, thus altering the local cultural ecology [49].

Missing from this hypothesized downward spiral is an initial perturbation that could result in an increase in cooperative norm violation (or perceived violation) in the neighborhood. Poverty and economic uncertainty are also striking differences between Neighborhoods A and B. Without middle class buffers of savings and credit, institutional safety nets, or strong reciprocal networks, crises such as illness create the potential for dire outcomes, thus altering the costs and benefits of defecting. For people already living at the margin, material crises might result in a higher probability of defection. Especially for crises that hit broad swaths of a community simultaneously, such as the widespread job loss in Neighborhood B resulting from the collapse of the shipbuilding and coal mining industries, one can imagine an increase in the frequency of defection that alters the descriptive cooperative norms enough to start a downward spiral in defection.

Importantly, although we hypothesize that poverty and economic uncertainty were linked to an initial perturbation of cooperative norm violation in the current study, the positive feedback of norm violation could continue in the absence of poverty. There has been

debate as to whether there are direct, as well as indirect, effects of poverty and/or income inequality on crime [50,51]. The story we have sketched is compatible with both possibilities, as an historical direct effect of poverty on norm violation may lead to cultural dynamics that persist beyond the duration of the poverty itself. (For a similar example of such cultural inertia, see [41], who argues that a transient change in the economics of crime can lead to persistently high crime rates, due to a postulated relationship between higher crime rates and decreased expectation of punishment.)

However, we can only speculate as to whether these dynamics are at play in Neighborhood B (outside of the 3PP game) and to what extent they can explain the observed high rates of crime and antisocial behavior. We acknowledge a number of limitations to our studies. We could not control the order at which participants looked at or filled out packet components. It is possible that participants 'justified' their behavior in the game with their questionnaire answers; however, we might then expect a robust positive effect of *difference £1 pound makes* on *theft*. Asking participants about norms could have positively influenced taking behavior in Neighborhood B by focusing participants on the high frequency of local cooperative norm violation [52]. Presenting Player 1s with the threat of punishment for theft could have decreased intrinsic motivation to behave cooperatively [53], although it is unclear how this would produce a spurious correlation between descriptive civic norms and theft in the game. We cannot account for the neighborhood residents who chose not to respond, although in both neighborhoods we likely reached a segment of the community biased towards prosocial preferences (registered voters and research participants). Finally, although participants were anonymous to each other in the game, they were not anonymous to us. The neighborhood differences in game behavior we observed could be partly attributed to

participants in Neighborhood A, but not B, regarding a university professor as someone in their social milieu and thus being concerned about reputational repercussions.

Conclusion.

We have two related suggestions for future study that may increase our understanding of why some communities appear to be stuck at uncooperative equilibria, despite concerted efforts by city planners to chart a different course [54], or even substantial temporal changes in the demographic makeup [10]. The first is further investigation of the potential for simultaneous multiple paths of positive feedback on cooperative norm violation, including not just conditional cooperation but also punitiveness and expectation of punishment. The second is consideration of how psychological adaptations for recognizing and adopting local norms, as well as biased in- and out-migration [47], can reinforce an antisocial culture.

Acknowledgements.

We thank residents of Newcastle Upon Tyne for participating in these studies, and we thank M.N. Grote, R. McElreath, and K. Rauch for helpful discussion and comments.

References.

1 Boyd, R. & Richerson, P. J. 1992 Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **195**, 171–195.

2 Fehr, E. & Gächter, S. 2000 Cooperation and punishment in Public Goods experiments. *Amer. Econ. Rev.* **90**, 980–994.

3 Herrmann, B., Thöni, C. & Gächter, S. 2008 Antisocial punishment across societies. *Science* **319**, 1362–7. (doi:10.1126/science.1153808)

4 Henrich, J. et al. 2006 Costly punishment across human societies. *Science* **312**, 1767–1770.

5 Wilson, D. S., O'Brien, D. T. & Sesma, A. 2009 Human prosociality from an evolutionary perspective: variation and correlations at a city-wide scale. *Evol. Hum. Behav.* **30**, 190–200. (doi:10.1016/j.evolhumbehav.2008.12.002)

6 Nettle, D., Colléony, A. & Cockerill, M. 2011 Variation in cooperative behaviour within a single city. *PloS One* **6**, e26922. (doi:10.1371/journal.pone.0026922)

7 Lamba, S. & Mace, R. 2011 Demography and ecology drive variation in cooperation across human populations. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14426–30. (doi:10.1073/pnas.1105186108)

8 Kocher, M., Martinsson, P. & Visser, M. 2012 Social background, cooperative behavior, and norm enforcement. *J. Econ. Behav. Organ.* **81**, 341–354. (doi:10.1016/j.jebo.2011.10.020)

9 Sampson, R. J. & Groves, W. B. 1989 Community structure and crime: testing social-disorganization theory. *Am. J. Sociol* **94**, 774–802. (doi:10.1086/229068)

10 Shaw, C. & McKay, H. 1942 *Juvenile Delinquency and Urban Areas*. Chicago, IL: University of Chicago Press.

11 Sampson, R. J., Raudenbush, S. W. & Earls, F. 1997 Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **277**, 918–924. (doi:10.1126/science.277.5328.918)

12 Bursik, R. & Grasmick, H. G. 1993 *Neighborhoods and Crime: The Dimensions of Effective Community Control*. New York: Macmillan, Inc.

13 Sampson, R. J., Morenoff, J. D. & Gannon-Rowley, T. 2002 Assessing "neighborhood effects": social processes and new directions in research. *Annu. Rev. Sociol.* **28**, 443–478. (doi:10.1146/annurev.soc.28.110601.141114)

14 Steenbeck, W. & Hipp, J. R. 2011 A longitudinal test of social disorganization theory: feedback effects among cohesion, social control, and disorder. *Criminology* **49**, 833–871. (doi:10.1111/j.1745-9125.2011.00241.x)

15 Fehr, E. & Fischbacher, U. 2004 Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87.

16 Camerer, C. F. & Fehr, E. 2004 Measuring social norms and preferences using experimental games: a guide for social scientists. In *Foundations of Human Sociality:*

Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies (eds J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Gintis), pp. 55–95. New York: Oxford University Press.

17 Cialdini, R. B., Reno, R. R. & Kallgren, C. A. 1990 A focus theory of normative conduct : Recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* **58**, 1015–1026. (doi:10.1037//0022-3514.58.6.1015)

18 Kelling, B. G. L. & Wilson, J. Q. 1982 Broken Windows. *The Atlantic Monthly* March, pp. 29–38.

19 Skogan, W. G. 1990 *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*. New York: Macmillan, Inc.

20 Keizer, K., Lindenberg, S. & Steg, L. 2008 The spreading of disorder. *Science* **322**, 1681–5.

21 Lochner, L. 2007 Individual perceptions of the criminal justice system. *Amer. Econ. Rev.* **91**, 444–460.

22 Knack, S. & Keefer, P. 1997 Does social capital have an economic payoff? A cross-country investigation. *J. Q. Econ.* **112**, 1251–1288.

23 Akaike, H. 1974 A new look at the statistical model identification. *IEEE. T. Automat. Contr.* **19**, 716–723.

24 R Development Core Team 2013 R: a language and environment for statistical computing. (Version 3.0.0, 2013; http://cran.r-project.org)

25 Venables, W. N. & Ripley, B. D. 2002 *Modern Applied Statistics with S*. Fourth. New York: Springer.

26 Mcelreath, R. 2012 rethinking: *Statistical Rethinking* book package. (R package version 1.10, https://github.com/rmcelreath?tab=repositories)

27 Ecklund, A. 2012 beeswarm: The bee swarm plot, an alternative to stripchart. (R package version 0.1.5, http://CRAN.R-project.org/package=beeswarm)

28 Wickham, H. 2009 ggplot2: elegant graphics for data analysis. New York: Springer.

29 Falk, A. & Fischbacher, U. 2002 “Crime” in the lab - detecting social interaction. *Euro. Econ. Rev.* **46**, 859–869. (doi:10.1016/S0014-2921(01)00220-3)

30 Gürerk, O., Irlenbusch, B. & Rockenbach, B. 2006 The competitive advantage of sanctioning institutions. *Science* **312**, 108–11. (doi:10.1126/science.1123633)

31 Boyd, R., Gintis, H. & Bowles, S. 2010 Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* **328**, 617–20. (doi:10.1126/science.1183665)

32 Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. 2003 The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3531–5. (doi:10.1073/pnas.0630443100)

33 Traxler, C. & Winter, J. 2012 Survey evidence on conditional norm enforcement. *Europ. J. Polit. Econ.* **28**, 390–398. (doi:10.1016/j.ejpoleco.2012.03.001)

34 Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J. & Griskevicius, V. 2008 Normative social influence is underdetected. *Pers. Soc. Psych. B.* **34**, 913–23. (doi:10.1177/0146167208316691)

35 Markowitz, F. E., Bellair, P. E., Liska, A. E. & Liu, J. 2001 Extending social disorganization theory: modeling the relationships between cohesion, disorder, and fear. *Criminology* **39**, 293–320.

36 Boyd, R. & Richerson, P. J. 1985 *Culture and the Evolutionary Process*. Chicago: The University of Chicago Press.

37 Sampson, R. J. & Raudenbush, S. W. 1999 Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *Am. J. Sociol.* **105**, 603–651.

38 Raihani, N. J. & Hart, T. 2010 Free-riders promote free-riding in a real-world setting. *Oikos* **119**, 1391–1393. (doi:10.1111/j.1600-0706.2010.18279.x)

39 Fischbacher, U., Gächter, S. & Fehr, E. 2001 Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett* .**71**, 397–404.

40 Irwin, K. & Simpson, B. 2013 Do descriptive norms solve social dilemmas? Conformity and contributions in collective action groups. *Soc. Forces* **91**, 1057–1084. (doi:10.1093/sf/sos196)

41 Sah, R. K. 1991 Social osmosis and patterns of crime. *J. Polit. Econ.* **99**, 1272–1295.

42 Glaeser, E. L., Sacerdote, B. & Scheinkman, J. A. 1996 Crime and social interactions. *J. Q. Econ.*, 507–548.

43 Wilson, D. S. & Csikszentmihalyi, M. 2007 Health and the ecology of altruism. In *Altruism and Health: Perspectives from Empirical Research* (ed S. G. Post), pp. 314–331. Oxford: Oxford University Press.

44 Henrich, J. & Gil-White, F. J. 2001 The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evol. Hum. Behav.* **22**, 165–196. (doi:10.1016/S1090-5138(00)00071-4)

45 Henrich, J. & Boyd, R. 1998 The evolution of conformist transmission and the emergence of between-group differences. *Evol. Hum. Behav.* **19**, 215–241.

46 Bowles, S. & Gintis, H. 2002 Social capital and community governance. *Econ. J.* **112**, F419–F436. (doi:10.1111/1468-0297.00077)

47 Chudek, M. & Henrich, J. 2011 Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends. Cogn. Sci.* **15**, 218–26. (doi:10.1016/j.tics.2011.03.003)

48 Akers, R. L. 2009 *Social Learning and Social Structure: A General Theory of Crime and Deviance*. New Jersey: Transaction Publishers.

49 Camerer, C. F. & Fehr, E. 2006 When does “Economic Man” dominate social behavior? *Science* **311**, 47–52.

50 Patterson, E. B. 1991 Poverty, income inequality, and community crime rates. *Criminology* **29**, 755–776. (doi:10.1111/j.1745-9125.1991.tb01087.x)

51 Bursik, R. J. & Grasmick, H. G. 1993 Economic deprivation and neighborhood crime rates, 1960-1980. *Law & Soc'y Rev.* **27**, 263–284.

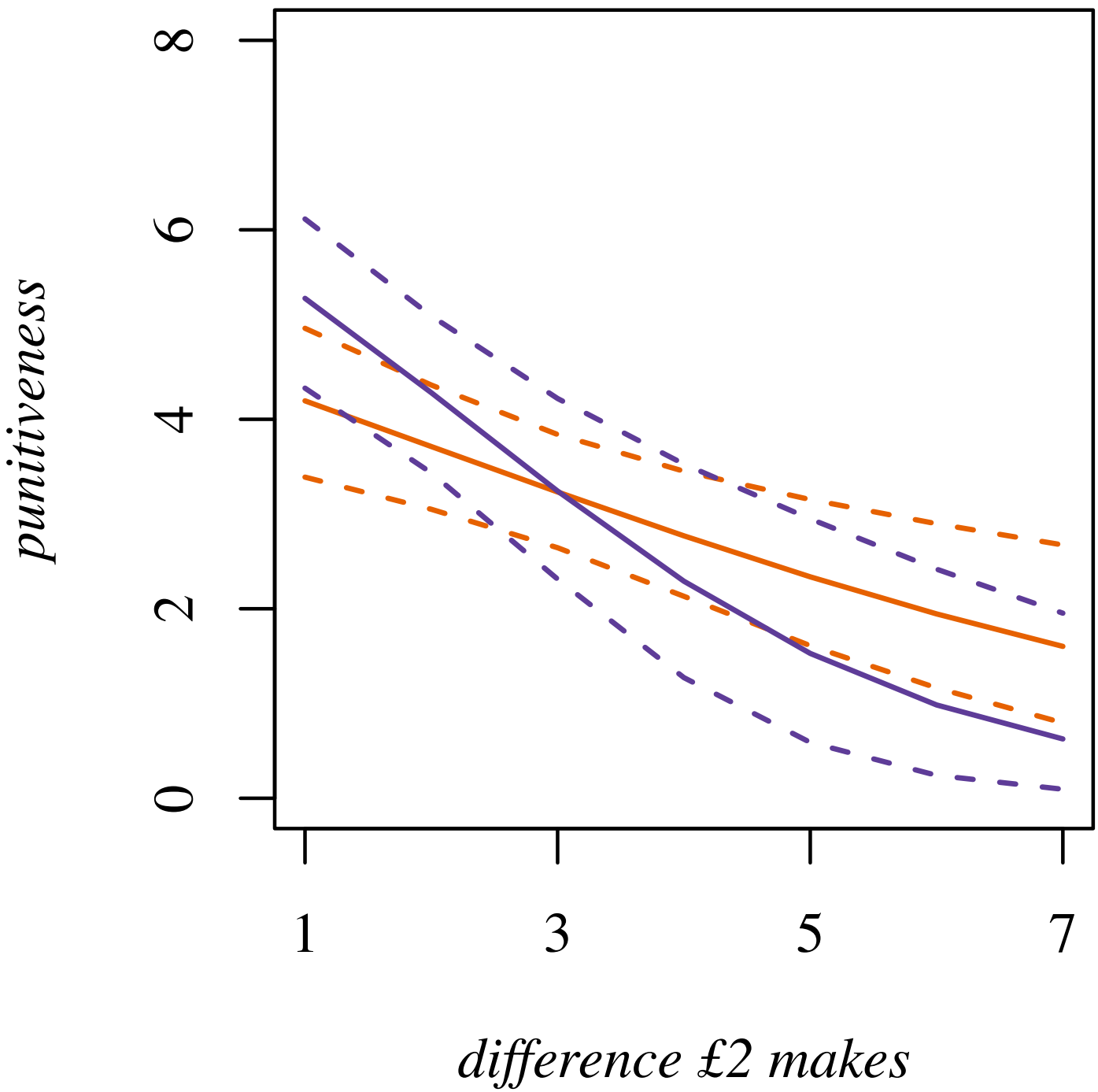
52 Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K. & Winter, P. L. 2006 Managing social norms for persuasive impact. *Social Influence* **1**, 3–15. (doi:10.1080/15534510500181459)

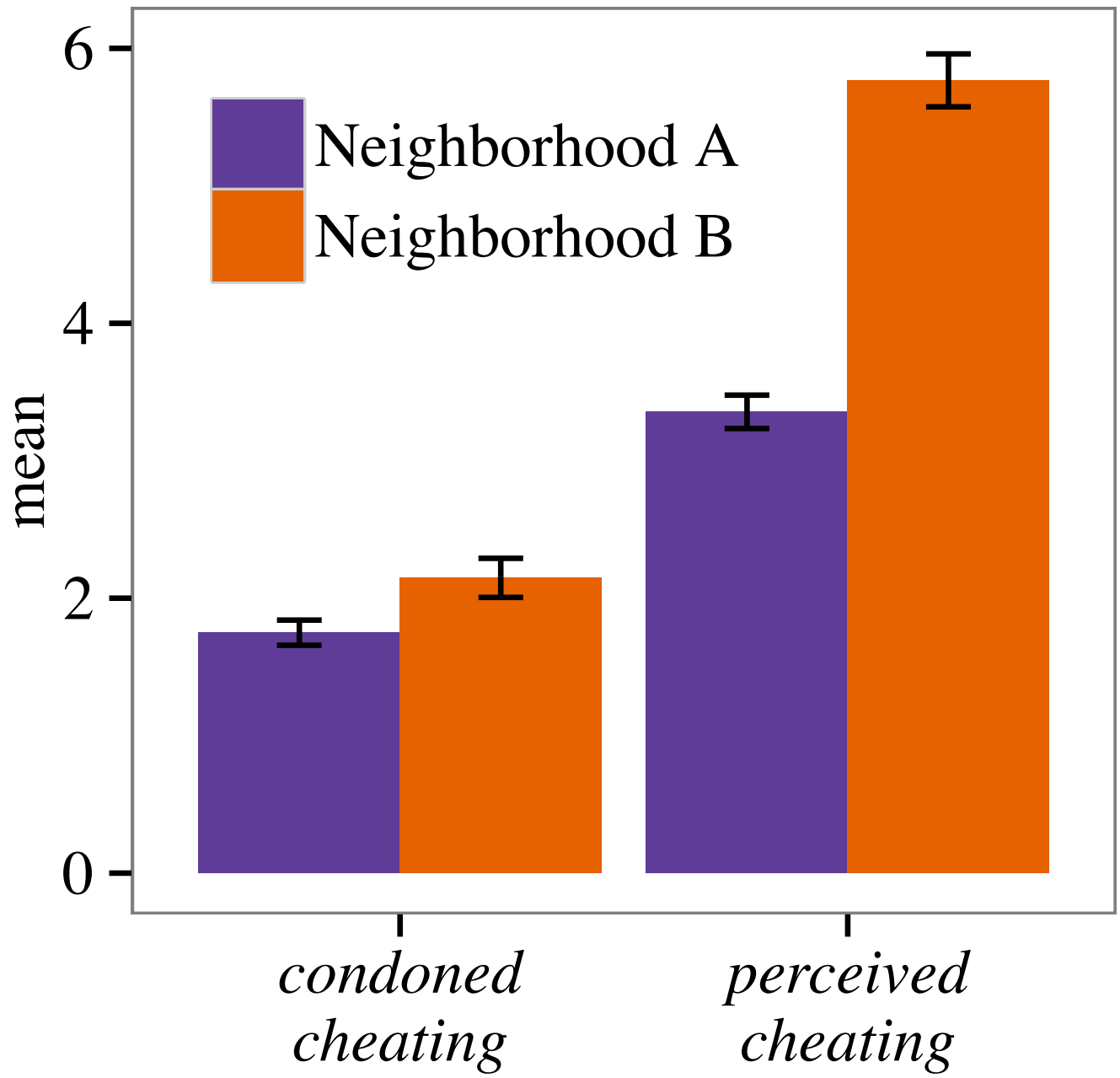
53 Frey, B. S. & Jegen, R. 2001 Motivation crowding theory. *J. Econ. Surv.* **15**, 589–611. (doi:10.1111/1467-6419.00150)

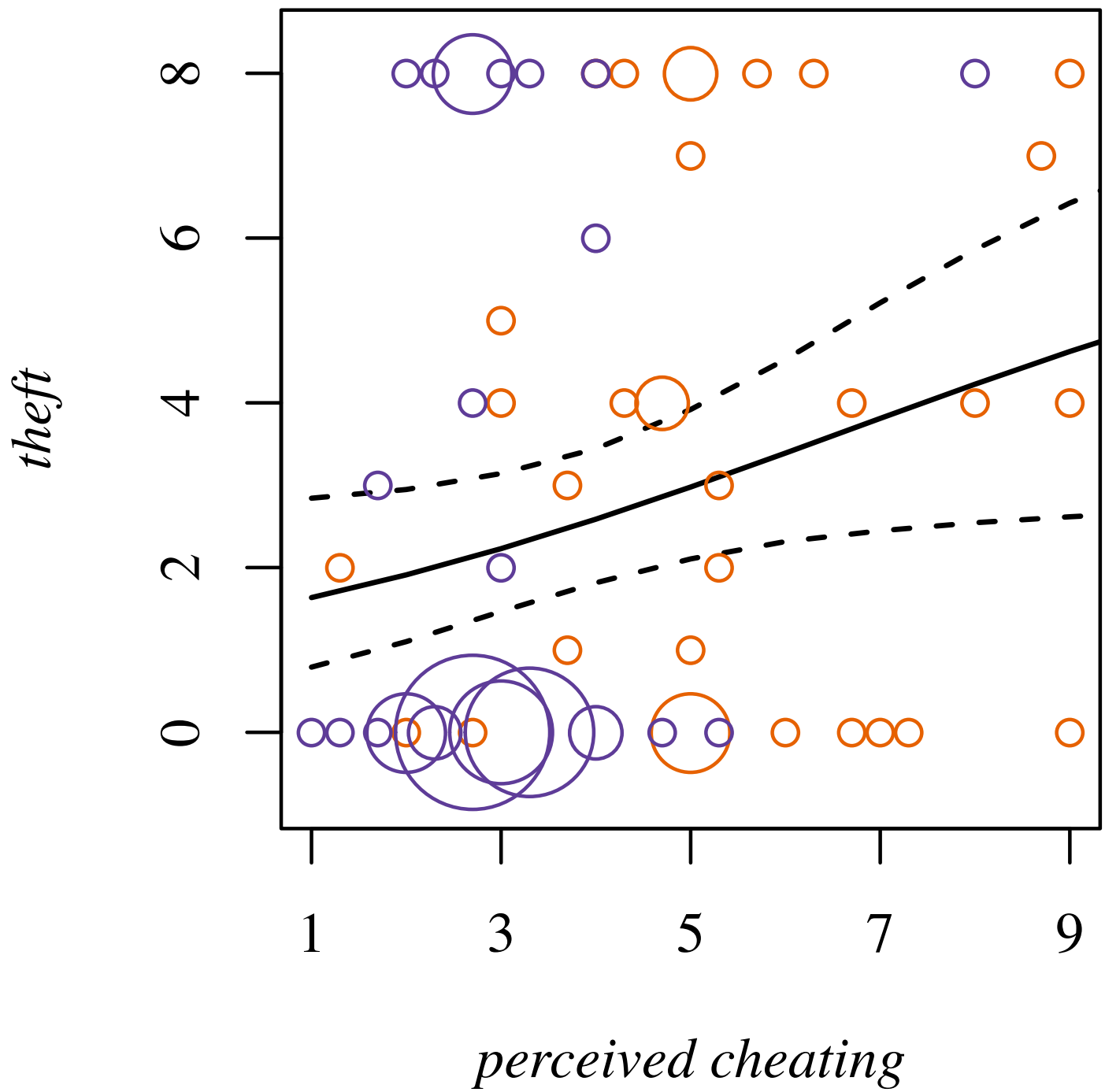
54 Robinson, F. 2005 Regenerating the West End of Newcastle: What went wrong? *Northern Economic Review* **36**, 15–41.

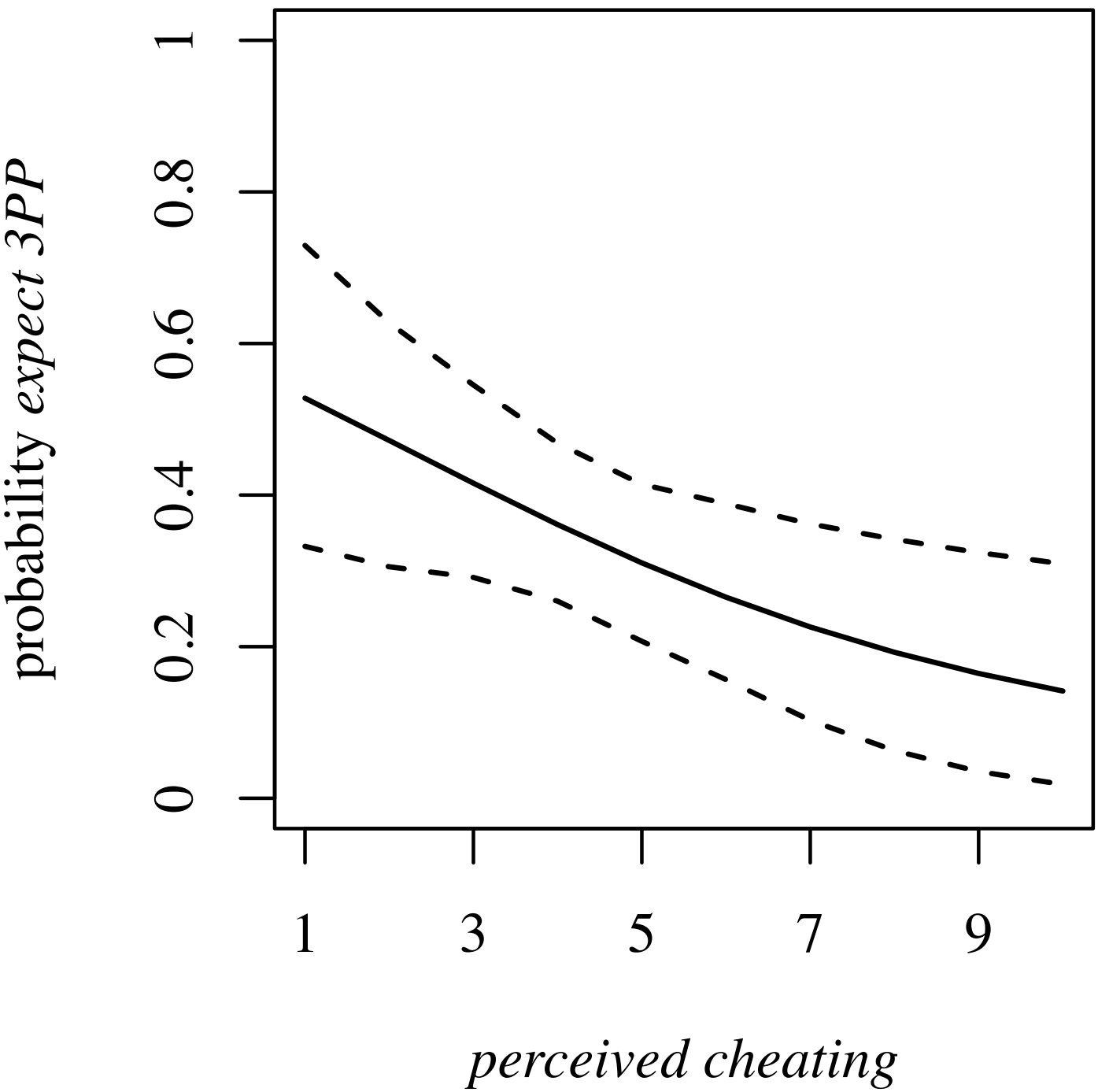
Figure legends.

1. Figure 1. Predicted *punitiveness* modeled as an interaction between *trust neighbors* and *difference 2 pounds makes*. Blue is 'high trust' (8; median *trust neighbors* score for Neighborhood A). Orange is 'low trust' (5; median *trust neighbors* score for Neighborhood B). Dotted lines are 95% confidence intervals.
2. Figure 2. Neighborhood means and standard errors for *condoned cheating* and *perceived cheating*. For *condoned cheating*, 1 = Never OK, 10 = Always OK, and for *perceived cheating*, 1 = No one would, 10 = Everyone would.
3. Figure 3. Predicted *theft* for Player 1. Dotted lines are 95% confidence intervals. Bubbles represent the actual data from Neighborhood A (blue) and B (orange). Size of the bubble corresponds to the number of observations.
4. Figure 4. Predicted probability of *expect 3PP* dependent on *perceived cheating*. Dotted lines are 95% confidence intervals.


punitiveness
0
2
4
6
8
1
3
5
7
difference £2 makes


6
4
2
0
mean
Neighborhood A
Neighborhood B
condoned cheating
perceived cheating


theft
0
2
4
6
8
1
3
5
7
9
perceived cheating


probability *expect 3PP*
0
0.2
0.4
0.6
0.8
1
1
3
5
7
9
perceived cheating