International Review of Law and Economics 37 (2014) 126–136

Contents lists available at ScienceDirect

International Review of Law and Economics

First impressions are more important than early intervention: Qualifying broken windows theory in the lab

Christoph Engel[a,*], Martin Beckenkamp[a], Andreas Glöckner[a,b], Bernd Irlenbusch[c], Heike Hennig-Schmidt[d], Sebastian Kube[a,d], Michael Kurschilgen[a], Alexander Morell[a], Andreas Nicklisch[e], Hans-Theo Normann[f], Emanuel Towfigh[a]

[a] Max Planck Institute for Research on Collective Goods, Bonn, Germany
[b] Georg-August-University, Institute for Psychology, Göttingen, Germany
[c] University of Cologne, Faculty of Management, Economics and Social Sciences, Germany
[d] Friedrich-Wilhelms-University, Faculty of Law and Economics, Bonn, Germany
[e] University of Hamburg, School of Business, Economics and Social Sciences, Germany
[f] Heinrich-Heine-University, Faculty of Business Administration and Economics, Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 25 May 2012
Received in revised form 20 June 2013
Accepted 15 July 2013

Keywords:
First impressions
Early vigilance
Social disorder and crime
Broken windows theory
Public good experiment

ABSTRACT

Broken windows: the metaphor has changed New York and Los Angeles. Yet it is far from undisputed whether the broken windows policy was causal for reducing crime. The scope of the theory is not confined to crime. The theory claims that crime is inextricably linked to social order more generally. In a series of lab experiments we put two components of this more general theory to the test. We show that first impressions and early punishment of antisocial behaviour are independently and jointly causal for cooperativeness. The effect of good first impressions and of early vigilance cannot be explained with, but adds to, participants' initial level of benevolence. Mere impression management is not strong enough to maintain cooperation. Cooperation stabilizes if good first impressions are combined with some risk of sanctions. Yet if we control for first impressions, early vigilance only has a small effect. The effect vanishes over time.

1. Motivation

Times Square, Manhattan, 1990: clearly not the place to be. You would have met all sorts of outcasts and would have exposed yourself to a serious risk of violent crime. Times Square, Manhattan, 2000: indulge in the world's most vibrant city, at its best. Don't be afraid of violence. The crime rate is substantially below the national average.[1] Usually Mayor Rudolph W. Giuliani and New York Police Dept. Commissioner William Bratton are credited with the success (Zimring, 2007). In recent years, William Bratton has repeated the New York success in Los Angeles (Wagers, 2008). In both cities, he explicitly relied on the "broken windows" policy (Kelling & Coles, 1996; Skogan, 1990; Sousa & Kelling, 2006; Wilson & Kelling, 1982).

In public perception broken windows theory is often equated with the abatement of crime. Yet this narrow reading misses the very point of the approach. The very essence of broken windows theory is the claim that crime is not by any means different from mere social disorder (see, e.g. Wilson & Kelling, 1982:5). If society does not care about social disorder, for minor that it may appear, it is on a slippery slope to ever and ever more severe forms of disorder and, eventually, crime. Crime is only the most manifest, and the socially most dreaded, expression of an effect the theory predicts if social order is visibly eroded, and not proactively restored. If society learns how to maintain social order more generally, by this very fact it keeps the risk of crime in check. In this paper, we exploit the generality of the theory and test two of its key components in a laboratory environment where social order is difficult to maintain. Specifically we test the following two claims of the theory: (1) depending on first impressions people make in an environment, they behave differently. Metaphorically speaking, the first broken window changes a neighbourhood. (2) If individuals quickly realize that their attempts at antisocial behaviour trigger a sanction, this tames antisocial behaviour. We expect that all debating the broken windows approach would want to know whether these implications of broken windows theory hold true.

The broken windows approach was inspired by an experiment conducted by Philip Zimbardo in 1969. Zimbardo simultaneously placed two otherwise identical cars in public spaces, one in the Bronx, the other in Palo Alto. Neither car had license plates, and the hood was open. Within 26 hours the car in the Bronx was totally

* Corresponding author at: Max-Planck-Institute for Research on Collective Goods, Kurt-Schumacher-Straße 10, D 53113 Bonn, Germany.
E-mail address: engel@coll.mpg.de (C. Engel).

[1] For details, see Uniform Crime Reports, at http://www.fbi.gov/ucr/ucr.htm.

0144-8188/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.irle.2013.07.005

pillaged and destroyed, while the Palo Alto car stayed pristine for an entire week. Once the experimenters themselves broke a window with a hammer, it went to ruins within hours, even in the sheltered and prosperous Californian town (Zimbardo, 1969).

Correlation analysis supports the claim that the broken windows policy, measured by the number of traffic tickets (Wilson & Boland, 1978), the number of arrests per police officer for disorderly conduct or driving under influence (Sampson & Cohen, 1988) or the number of misdemeanour arrests (Corman & Mocan, 2005; Kelling & Sousa, 2001), contributed to the decline in serious crimes, even if one controls for economic conditions and for crime deterrence (Corman & Mocan, 2005; see also Cruz Melendez, 2006 for the link to the "Moving to Opportunity" Program). Along the same lines, time series evidence from Switzerland shows tougher enforcement of mild crimes to reduce the incidence of severe crimes in later years (Funk & Kugler, 2003). In Los Angeles, neighbourhood deterioration preceded the onset of crime rates (Schuerman & Kobrin, 1986). Yet, other studies did not find a significant effect (Geller, 2007; Katz, Webb, & Schaefer, 2001; Novak, Hartman, Holsinger, & Turner, 1999). They used a complex index of perceived social disorder as the independent variable (Sampson & Raudenbush, 1999). Information about law-abiding or the number of abandoned buildings did not have a significant influence either on young males' beliefs about the risk of being convicted (Lochner, 2007); (see also the mixed results by Rosenfeld, Fornango, & Rengifo, 2007; Taylor, 2001) (further see Blumstein, 1995; Bowling, 1999; Messner et al., 2007: on the link to the exogenous evolution of the drug market). Yet others argue that the broken windows approach should be embedded into a broader assessment of the relationship between neighbourhood change and crime (Fagan, 2008; Taub, Taylor, & Dunham, 1984). Most importantly, it is far from undisputed whether correlation can be interpreted as causation (Harcourt, 1998, 2001, 2005; Harcourt & Ludwig, 2006; Karmen, 2000; Sampson, Morenoff, & Gannon-Rowley, 2002).

In the field, the fact that the window is not fixed (that panhandlers are free to molest passers by; that drunks congregate in the park; that rowdies menace shopkeepers) also gives a signal to those who have always been living in the area. They may read this as evidence that social cohesion is eroding. Yet normally they have many more sources of information, from which they draw their personal conclusions. They talk to each other, they read the local newspaper, they address themselves to the authorities. Therefore, in the field the effect of the signal is hard to identify (cf. Fagan, 2008:109 f. on identification problems when estimating the relationship between neighbourhood change and crime). Equally hard is identifying the motives of those who seem to behave differently. Do they move to another neighbourhood simply because they can afford it, because they want to send their children to a better school, because a new street has brought another suburb within reach – or do they move out to protect themselves from the perceived risk of crime? Is the city centre less populated because people prefer to meet in private clubs, because shopping malls in the outskirts attract customers, because people spend more time watching TV – or because they infer from the (real or metaphorical) broken windows that the centre is no longer safe?

To avoid such identification problems, in the experiments reported in this paper we create an artificial neighbourhood. The experimental setting exposes participants to a social dilemma. Individually, each participant is best off if the remaining group members contribute to a joint project while she freerides. Participants interact in a randomly composed group of four over ten announced periods. This design gives us a clean measure of (anti-) social behaviour. The less a participant contributes, the more she imposes damage on the remaining group members.

For our first research question, the explanatory variable of interest is the impression participants happen to gather in the first period. We operationalize this as the mean contribution by the remaining three group members, in the first period. We measure the causal effect of first impressions on contributions in later rounds. First impressions do indeed have strong explanatory power. The effect does not collapse with participants' idiosyncratic social value orientation, as expressed in participants' own contribution to the public project in the first round of interaction, i.e. while they are unaware of the cooperativeness of the remaining members of their group. The average amount the remaining group members have contributed in the first round explains their choices until the penultimate round; in the final round, selfishness wins the day, even with participants who were willing to support the joint project in earlier periods. The effect of first impressions does not disappear if we control for learning, as expressed in an individual's contribution in the previous round. The effect is visible for participants who have contributed more, and for those who have contributed less than the average of their groups in the first period. It thus is not confined to those strongly, or to those little socially minded.

Broken windows theory has been heavily used in criminal policy, as a motivation for and justification of zero tolerance with respect to petty crime. One should therefore expect that would-be offenders are more likely to desist from antisocial behaviour if they are deterred. One could further expect that community members are willing to police disorder themselves if given the opportunity, but that they are less likely to do so if they have reason to fear for revenge. This is essentially what we find. If participants are able to express disapproval and deter freeriding through costly punishment, with sufficiently favourable first impressions cooperation is stabilized in the long run, even if those punished are given a chance to strike back. If sanctions are excluded by design, cooperation decays. But conditional on first impressions, average contributions are higher, and the decay is slower.

For our second research question, the explanatory variable is reactions to antisocial behaviour in the first round of interaction. If we control for first impressions, the effect is small in early rounds, and becomes insignificant in later rounds. The critical cause is first impressions, not early vigilance. This is an important piece of news for the policy debate. In public perception, broken windows policies have been associated with being tough on crime, and on petty crime and disorder short of criminal infraction more specifically. Our data suggest that this is at most a secondary cause. If freeriders realize that crime and disorder have consequences, they behave better. This, in turn, gives others a better impression of the kind of behaviour to be accepted in this society. These impressions are key, not punishment per se.

Experiments of necessity pay a price for control. They have to abstract from many features of the real life phenomenon they aim to explain. Our experiment is no exception. We abstract from the possibility that perceived disorder attracts criminals to a community who did not inhabit it before. We are not studying the sudden change of a previously orderly neighbourhood to the worse, but have everybody start from scratch in a new environment. In our setting, disorder and crime are only distinct by the degree of antisocial behaviour, and are not qualitatively different. Loyal participants may at most fear losing some of their experimental income, not their lives, health or belongings. Despite all these simplifications, we believe the price for experimental control to be affordable.

The closest analogue in the field is the behaviour of those who newly arrive in a neighbourhood, be that a family who moves in, a child who goes to a new school, or a person who visits a new area. That way, our results also speak to the class of persons broken windows theory is most interested in: criminals who consider entering a community since, reading the signals, they believe they stand a fair chance to get away with their illegal acts.

In our experiment, there is no formal separation between disorder and crime. But through the gradual nature of our dependent variable, we have a good proxy for "criminal invasion" (Wilson & Kelling, 1982): if some have been a little below others' expectations initially, chances are others will freeride even more intensely in later periods. This is exactly how contributions decay in groups where first impressions have not been good.

In our experiments, just a few coins are at stake. In the field, the inhabitant of a neighbourhood in decay may have to leave the house in which she was born, she may see her property burglarized, and may even fear for her life. In the field, through the power of fear small initial disorder may easily start a vicious cycle. One such story might be: initial signs of disorder cause fear. Residents stay at home. This weakens social control. First offenders invade the neighbourhood. Even more residents refrain from actively maintaining order. Serious criminal activity is pulled to the neighbourhood. Yet this makes it all the more noteworthy that, in our much less dramatic setting, we also find a strong and lasting effect of first impressions.

Seemingly, the problem of criminal policy is different in that the focus is not on proactive contributions to a common good, but on the absence of antisocial behaviour. Yet as a group, the inhabitants of an area are best off if everybody's integrity and property are respected, while individually, a criminal is best off if only the others refrain from crime, and she finds ample prey. The dilemma even has a second level (cf. Heckathorn, 1989; Yamagishi, 1986). Individually, each member of the community is best off if others bear the cost of policing order, while she enjoys the peaceful environment. From the perspective of broken windows theory, this is not a minor issue. In their programmatic article, Wilson and Kelling claim: "The essence of the police role is to reinforce the informal control mechanisms of the community itself" (Wilson & Kelling, 1982:6).

In other respects, our experiments exactly capture the mechanism adherents of broken windows theory believe to be crucial. In our experimental groups, all rule-making is implicit and local, as are sanctions. The communities have to rely on the self-policing of vague rules of conduct (Wilson & Kelling, 1982).

The remainder of this paper is organized as follows. Section 2 links our work to the related literature. Section 3 describes the dataset and the experimental designs. Section 4 presents and analyses the results. Section 5 discusses implications for broken windows theory.

2. Related literature

The closest analogue to our study in the legal literature is a field experiment that randomly exposed 12 of 24 matched violent crime places in Jersey City to intense police scrutiny and intervention. In the places chosen, crime rates dropped substantially, while they did not in the unaffected places (Braga et al., 1999). A further careful field experiment randomly exposed crime and disorder hot spots in Lowell, Mass. to "shallow" vs. intense police efforts to restore order, to show that situational prevention strategies were most effective in curbing crime (Braga & Bond, 2008). In a similar vein, in a series of sociological field experiments, when there were signs of disorder, like graffiti, abandoned shopping carts, or bicycles locked where they were not supposed to be, this induced passers-by also to break these and other rules (Keizer, Lindenberg, & Steg, 2008).

Our dataset differs from all these studies in that our "intervention" is much more light-handed; it is confined to the first impressions subjects happen to make and, if the design allows for that, to receiving punishment in the first round of interaction. Since we conducted lab experiments, we need not have second thoughts about the influence of explanatory variables beyond our control. A further advantage of our approach stems from the nature of both the dependent and the independent variables. In the field, both are categorical: people either break the law or they obey it; people either see disorder or they do not. In our setting, "disorder" is measured by the distance from socially optimal behaviour, and socially desirable behaviour is measured by the amount bystanders contribute to the joint project. Likewise, we not only observe that a participant is punished, but also how severely. We are able to distinguish between the overall level of disorder and the maximum disorder participants experience in the group of which they happen to be a member. Since all our data is from games repeated over 10 periods, we can analyse the dynamics triggered by favourable or unfavourable first impressions, and we can check when a beneficial effect of first impressions or early punishment vanishes.

In the economics literature, the closest analogue is an experiment where, in a first stage, participants were screened for their cooperativeness. In the second stage, they played a standard public-good game, knowing they were interacting with partners that scored like them in the pre-test. In a voluntary contribution mechanism, this unequivocally increased cooperation, even for those scoring low in the pre-test. However with punishment, overall contributions decayed, due to very poor performance of those scoring low in the pre-test (Gächter & Thöni, 2007). The effect of sorting is positive throughout if subjects are rematched every round according to their cooperativeness in the previous round (Gunnthorsdotir, Houser, & McCabe, 2007). Likewise, if groups have a chance to exclude freeriders, this improves cooperation in a dilemma setting (Cinyabuguma, Page, & Putterman, 2005; Croson, Fatas, & Neugebauer, 2008), as does a mechanism that allows members to self-select into groups (Page, Putterman, & Unel, 2005), in particular if freeriders are effectively excluded by a rule that sacrifices a portion of the group income to outsiders (the Red Cross, as it was) (Brekke, Hauge, Lind, & Nyborg, 2009). Our study differs from this literature in that all we use is an element present in any public good game, and in any real life social dilemma: the first impressions participants happen to make, and the experience of vigilance.

Our paper is also related to the literature on conditional cooperation. The classic experimental tool to make this motive visible is Fischbacher, Gächter, and Fehr (2001). Using the strategy method, they show that a large fraction of participants condition their own contributions to a linear public good on contributions of other group members. In a follow-up experiment, they also explain contributions by others' contributions in the previous periods, and by beliefs (Fischbacher & Gächter, 2010). This is in line with the data analysis strategy used by Keser and van Winden (2000). Others have also found significant effects of the minimum contribution of others in the previous period (Croson, Fatas, & Neugebauer, 2005) and of the individual's own contribution in the previous period (Ashley, Ball, & Eckel, 2010). In field experiments on charitable giving, information about how much others have given in similar situations also had a significant effect (Croson & Shang, 2008; Frey & Meier, 2004). Our contribution is to show how little it takes to trigger these effects, and how long the effects last. Specifically, we do not explain this period's choices by experiences from the previous period, but from the very first moment when interaction has started. We show that these experiences matter until the penultimate period, if there is punishment even until the final period, and that these first impressions are much more important than early vigilance, i.e. punishment in the starting period.

Finally, we make a methodological contribution to the burgeoning field of experimental criminology (Farrington, 2003, 2006; Farrington & Welsh, 2005; Telep, 2009). We show how meaningful and productive it is to apply standard tools from experimental economics to a longstanding issue in criminology.

3. Design and data

In our experiments, we expose participants to a dilemma. Players interact repeatedly for 10 periods in groups of size four. The situation is fully symmetric, which all participants know. Specifically each player has the following payoff function π_i:

$$\pi_i = 20 - g_i + 0.4 * \sum_{k=1}^{4} g_k$$

Thus each period each participant receives 20 tokens from the experimenter. She is free to keep all of them, or to invest them partly or fully in the joint project. Each token she keeps gives her 1 token. Each token she invests only gives her 0.4 tokens. Yet she also receives 0.4 tokens for every token any other group member has invested into the project. Hence the entire group gains 1.6 tokens from each token invested. A participant is best off if all others have contributed fully, while she has contributed nothing. She then has 20 – 0 + 0.4*60 = 44 tokens. She is worse off if all others have contributed nothing while she alone has invested fully. She then has 20 – 20 + 0.4*20 = 8 tokens. If all contribute their entire endowments, all have 20 – 20 + 0.4*80 = 32 tokens. If all keep their entire endowments, all have 20 – 0 + 0 = 20 tokens.

In the literature, an experimental game with this structure is called a voluntary contribution mechanism (VCM). Our dataset also encompasses data from two variants. In the first variant, after all group members have decided how much to contribute to the project, they are informed about contributions by the remaining three group members. They are given the opportunity to react by spending some of their period income on reducing other group members' incomes. In the second variant, after participants have decided about punishment, players receive feedback about the punishment decision made by others and can then spend some of the remaining period income to punish those who have punished them. Since we wanted to merge our own data with data from other experimenters, we have kept the non-linear punishment technology originally used by Fehr & Gächter (2000). It is explained in the Appendix.

Public goods experiments are a standard tool of experimental economics. In our own experiments, we moreover have used parameters that are standard in this literature. This provides us with the opportunity to test the effect of first impressions and of early vigilance in a much larger dataset. To that end, the following is partly a reanalysis of data from public good experiments that are already published (Denant-Boèment, Masclet, & Noussair, 2007; Herrmann, Thöni, & Gächter, 2008; Nikiforakis, 2008), and partly of our own, hitherto unpublished data. The total dataset comprises 15,320 datapoints, or data from 1532 participants. Table 1 informs about the different design features and parameters in more detail. All games are played in groups of four, with an endowment of 20 tokens per player. Each token contributed to the project increased each group member's payoff by 0.4 tokens.

The first column indicates whether participants had no technology for targeted sanctions (VCM), or whether they could punish each other without (Pun) or with the risk of counterpunishment (CPun). The second column indicates the origin of the data, where *MPI* denotes our own experimental data, *DEN* is data provided by Denant-Boèment et al. (2007),[2] *NIK* is data taken from Nikiforakis (2008), and *HER* is data published in Herrmann et al. (2008), which consists of 16 structurally identical experiments run in different countries.[3] The third column gives the total number of individual decisions in the respective dataset. More detail on experimental procedure and on the instructions of our own, new data is to be found in the Appendix.

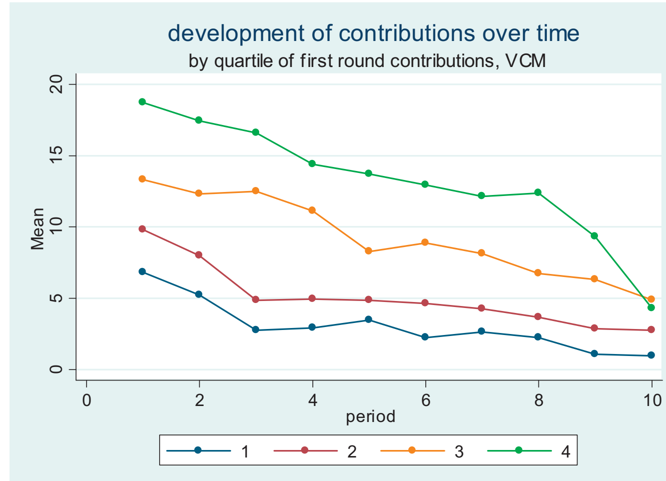

Fig. 1. First impressions in an institution free environment. dv: average group contribution to public good. Groups are classified by average contributions in the first round.

The fourth and fifth columns denote which punishment or, as the case may be, counter-punishment technologies were used. Here, *1:3* indicates that a linear technology was used where each punishment point assigned costs one token and reduces the other's payoff by three tokens, *FG* indicates that the non-linear technology introduced by Fehr and Gächter (2000) was used, which is described in the Appendix. *SEV* indicates that a severe technology was used, where each assigned counter-punishment point costs one token and reduces the receiver's net payoff (after the effect of received and the cost of given punishment are subtracted) by 25%. The last column describes the amount of information that subjects were given on the counter-punishment stage, where *own* indicates that subjects only knew the amount of punishment they had received themselves, *others* indicates that subjects only knew by how much the other members of the group had been punished, and *all* indicates that subjects knew whether and by how much each subject had been punished.

4. Results

We have two independent variables: first impressions and early vigilance. We address them in turn.

4.1. First impressions

We first consider the effect of first impressions in an environment where targeted reactions to freeriding are not possible, i.e. in a voluntary contribution mechanism. Fig. 1 demonstrates that the willingness to behave in a socially responsible manner strongly depends on first impressions. If the group mean was low in the first round, contributions stay very low. The higher mean contributions in the first round, the higher they are later. Eventually, contributions decay. Even excellent first impressions cannot remedy the absence of any institutional safeguard against freeriding. Yet

[2] The original dataset of Denant-Boèment et al. (2007) contains 20 periods. To keep datasets comparable, only the first ten periods of each matching group are considered in our analysis.

[3] Athens (number of observations N = 440), Bonn (600), Boston (560), Chengdu (960), Copenhagen (680), Dnipropetrovs'k (440), Istanbul (640), Melbourne (400), Minsk (680), Muscat (520), Nottingham (560), Riyadh (480), Samara (720), Seoul (840), St. Gallen (960), Zurich (920).

Table 1
Data structure.

Game-type	Dataset	# obs.	Punishment technology	Counterpunishment Technology	Punishment feedback
VCM	MPI	240	–	–	–
VCM	NIK	960	–	–	–
VCM	MPI	400	–	–	–
Pun	DEN	240	FG	–	–
Pun	MPI	240	FG	–	–
Pun	NIK	480	FG	–	–
Pun	HER	10,400	1:3	–	–
CPun	MPI	680	FG	FG	Own
CPun	NIK	480	FG	FG	Own
CPun	DEN	240	FG	FG	All
CPun	DEN	240	FG	FG	Others
CPun	DEN	240	FG	FG	Own
CPun	MPI	480	FG	SEV	Own

differences in first impressions remain visible until the end of the game.

As Table 2 shows, the visual impression is fully borne out by statistical analysis.[4] Critically, these regressions control for individuals' own contribution in the first round. It has strong explanatory power for contributions in later rounds. But even conditional on the idiosyncratic level of cooperativeness, we find a strong effect of the average contributions of the remaining group members in the first round. Actually in models 1 and 2 the latter coefficient is even larger. This suggests that first impressions are even more important than an individual's own social value orientation. Models 2 and 3 interact first impressions with the time trend.[5]

One may wonder whether the significant effect of first impressions is driven by choices in early periods. Arguably, first impressions only influence first reactions, and all the rest is a result of group dynamics. Fig. 2 shows that this is not the case. This figure compresses the results from nine separate regressions. From regression to regression we reduce the sample by another period, and only consider data from that period on. Except if we test the final period in isolation, for all periods do we find a strong, and a strongly significant effect of first impressions. Even if we only analyse data from a few final periods, the effect of what a participant has seen in the first period remains almost as strong as in the second period.

Another competing explanation is learning. If it was true, the effect of first impressions should become insignificant once we control for a participant's contribution decision in the previous period. First impressions would translate into second round choices; second round choices would translate into third round choices, and so forth. Last round choices would have to contain all the information necessary for predicting this period's choices. As Fig. 3 shows, this alternative interpretation is not correct. Again with the exception of the final period, we find a significant effect of first impressions if

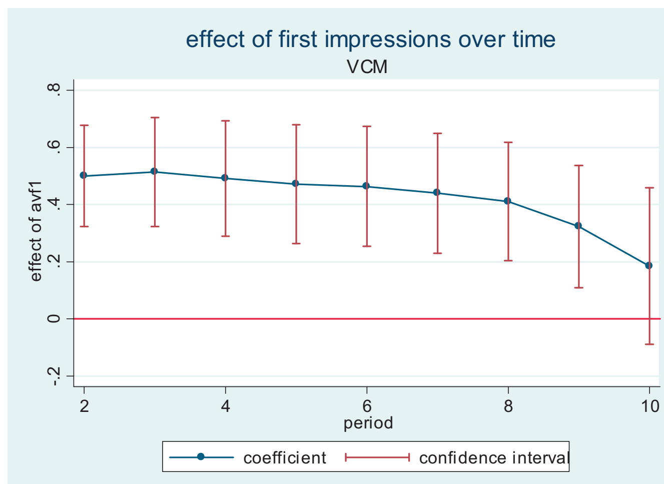

Fig. 2. Long lasting effect of first impressions. Coefficient from model 1 of Table 2 for a series of regressions considering only data from the indicated period on model for period 10 only is OLS (since the data is no longer panel data) with standard errors clustered for groups.

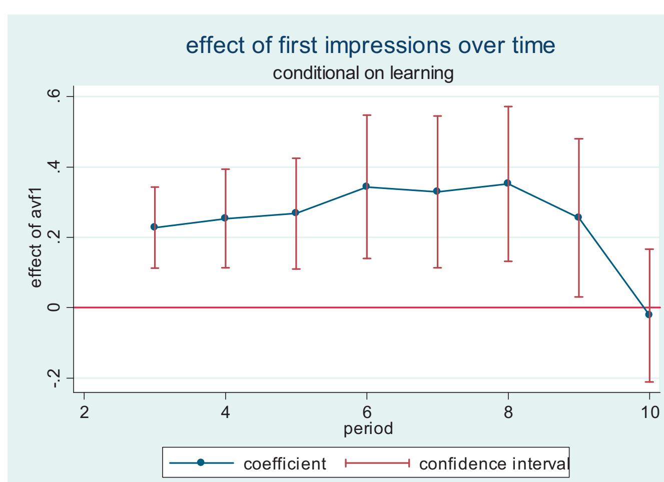

Fig. 3. First impressions vs. learning. Arrelano Bond systems estimator, one lag, robust standard errors clustered for groups. Coefficient from a series of regressions considering only data from the indicated period on. Model for period 10 only is OLS (since the data is no longer panel data) with standard errors clustered for groups.

[4] From each participant, we observe 10 contribution choices. Each individual stays a member of her group of 4 for the entire experiment. This gives us nested data. We match the data generating process by a mixed effects model. We thus estimate a random effect for groups, and another random effect for individuals nested in groups, plus residual error.

[5] Model 3 is most informative. The model predicts that contributions decay rapidly if both this individual and the remaining group members have kept their entire endowment in the first period. The more the other group members have contributed in the first round, the more contributions are stable. The model predicts that there is no decay if the remaining group members have on average contributed 16.64 tokens (2.247/.135 = 16.644). Over time, the decay flattens (the quadratic time trend is positive). But this effect is most pronounced if the average contributions of others have been low in the first round (interaction between average contributions of others and the quadratic trend).

Table 2
First impressions in an institution free environment.

	Model 1	Model 2	Model 3
Individual contribution in period 1	.365***	.365***	.365***
Average contribution of the remaining group members in period 1	.500***	.785***	.339*
Period	−.798***	−.220+	−2.247**
Period^2			.169**
avf1*period		−.048***	.135**
avf1*period^2			−.015***
cons	1.398	−2.073	2.882
N	1440	1440	1400
p model	<.001	<.001	<.001

Linear mixed effects, choices nested in individuals nested in groups.
Data from periods 2-10.
avf1: average contribution of the remaining group members in period 1.
Hausman test insignificant on all models.
* $p < .05$; ** $p < .01$; *** $p < .001$; + $p < .1$.

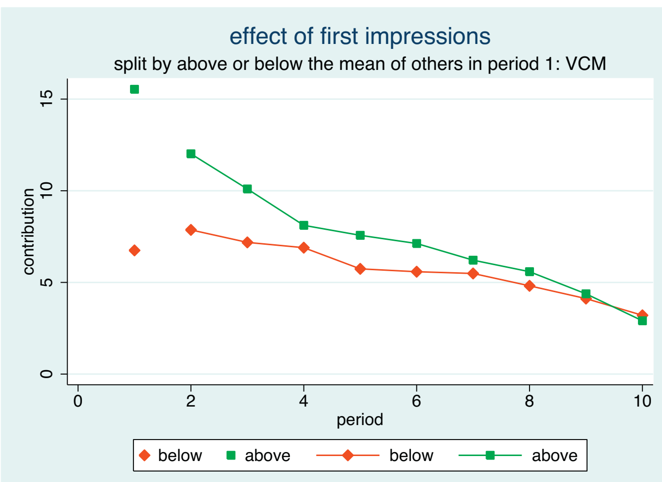

Fig. 4. First impressions by relative position in the first round.

we control for learning.[6] There are two forces that simultaneously affect current choices: first impressions *and* learning.

One may further wonder whether the effect of first impressions is confined to particularly selfish, or to particularly socially minded, participants, or whether, at least, it plays itself out differently for both groups. Fig. 4 shows that both subgroups directly adjust to what they have seen in the first period, and then quickly converge. From period 4 on, their behaviour becomes practically undistinguishable.

The regression of Table 3 supports the visual impression. Neither the two-way interaction between the average contribution of the remaining group members in period 1 and whether this participant was above or below this benchmark, nor the three-way interaction of the former with the time trend, are significant. The effect of first impressions is not different for those who are more from those who are less socially minded than their peers. The effect of first impressions does also not play itself out differently over time for these two subgroups. For consistency with Table 2, we also report a model that includes period^2 and interaction terms. We lose the interaction terms between first round impressions and time, and between having been above average of others in period 1 and time. Yet the message remains unchanged: Neither the two-way interaction between the average contribution of the remaining group members in period 1 and whether this participant was above or below this benchmark, nor the three-way interaction of the former with the time trend, are significant.

Finally, the effect of first impressions might be conditional on the homogeneity of groups. Arguably, the overall effect might just reflect that enough groups have been sufficiently homogeneous in the first place. Table 4 shows that this is not the case. If we additionally condition choices in later rounds on the minimum contribution of one of the remaining group members, the effect of the average contribution of the remaining group members becomes even stronger. Since we control for the average contribution, a small minimum contribution directly captures the spread of contributions, i.e. homogeneity. Again for consistency with Tables 2 and 3 we further control for interaction between average and minimum contributions and time, and in the final model we further control for non-linear developments over time plus their interaction with first impressions. With these additional controls, we do indeed find some additional effects that are conditional on local circumstances. The beneficial effect of good first impressions (avf1) becomes a little less important over time (models 2 and 3), but the net effect remains strongly positive and significantly different from 0 even in

Table 3
First impressions by relative position in the first round.

	Model 1	Model 2
Average contribution of the remaining group members in period 1	.858***	.476*
Above average of others in period 1	9.706***	11.368**
avf1*above	−.249	−.381
Period	.454*	−1.199
avf1*period	−.078***	.079
Above*period	−.856**	−1.536
avf1*above*period	.019	.073
Period^2		.138
avf1*period^2		−.103*
Above*period^2		.057
avf1*above*period^2		−.004
cons	−1.831	2.211
N	1440	1440
p model	<.001	<.001

Linear mixed effects, choices nested in individuals nested in groups.
Data from periods 2-10.
avf1: average contribution of the remaining group members in period 1.
Above: dummy that is 1 if the contribution of this individual was above the average contribution of the remaining group members in period 1.
Hausman test insignificant.
* $p < .05$; ** $p < .01$; *** $p < .001$; + $p < .1$.

[6] Technically, we estimate a dynamic panel with a one-period lag. Such models are known to be inconsistent, which is why we must instrument. If we use the original Arellano–Bond estimator, the time-invariant effect of first impressions drops out. We therefore use the systems estimator, i.e. a method of moments approach. There is no mixed effects version of this estimator. We capture the dependence of observations at the group level by clustering standard errors. Again we estimate a series of eight regression (not nine regressions since the Arellano–Bond estimator uses two lags), and reduce the sample by one period from regression to regression.

Table 4
First impressions conditional on local heterogeneity.

	Model 1	Model 2	Model 3	Model 4
Individual contribution in period 1	.386***	.386***	.386***	.386***
Average contribution of the remaining group members in period 1	.697***	.982***	1.037***	1.206***
Minimum contribution of the remaining group members in period 1	−.161+	−.161+	−.204+	−.685**
Period	−.798***	−.220+	−.158	−.504
avf1*period		−.048***	−.057**	−.126
minf1*period			.007	.204**
Period^2				.029
avf1*period^2				.006
minf1*period^2				−.016**
cons	−.131	−3.602+	−3.971+	−3.127
N	1440	1440	1440	1440
p model	<.001	<.001	<.001	<.001

Linear mixed effects, choices nested in individuals nested in groups.
Data from periods 2-10.
Hausman test insignificant.
* $p<.05$; ** $p<.01$; *** $p<.001$; + $p<.1$.

the final period.[7] Likewise, in model 4 the net effect of the minimum contribution to the public good in the first round (minf1) remains significantly negative even in the final period.[8]

We conclude

Result 1. *In a linear public good, average contributions of the remaining group members in the first round determine contributions in later rounds. This holds irrespective of the individual degree of cooperativeness. The effect lasts until the endgame effect kicks in. The effect does not collapse with learning. It is not confined to those initially above or below the average. It is not conditional on initial homogeneity*.

Result 1 is important for criminal policy as a counterfactual. What is to be expected if all those who care cannot react to a perceived deterioration of socially desirable behaviour, and to a vicious cycle of freeriding? Even in such an institution poor environment, first impressions have a strong, and a strongly beneficial effect. Yet they are not strong enough to stop the gradual decay of socially desirable behaviour. As Fig. 1 demonstrates, groups that were good in the beginning remain better than groups that had more freeriding at the outset. Whether windows are broken matters, even absent institutional intervention. Yet eventually even the best groups observe the decay of socially minded behaviour. In the long run, the implicit norm is less and less obeyed.

Out there in the field, broken windows can be repaired. If someone is observed breaking a window, she may attract a reaction by bystanders. Yet intervention may be risky. Those who have pleasure from disturbing order might react aggressively against acts of vigilance. Fig. 5 shows that giving those who dislike freeriding a chance to react matters strongly. This also holds if those punished have a chance to strike back. Critically for our research question, in both environments we find a strong effect of first impressions.

Table 5 provides statistical support. In both institutional environments, first impressions again have a strong, and a highly significant beneficial effect on contributions in later rounds. Using the same tests as with a voluntary contribution mechanism, we can show that, both with punishment and with counterpunishment, the effect of first impressions is long lasting; that it does not collapse with learning; that it is not confined to particularly socially minded individuals; that it is not confined to homogeneous groups. To save space, we do not report these results in detail,[9] and confine ourselves to stating

Table 5
Effect of first impressions in richer institutional environments.

	Punishment	Punishment and counterpunishment
Individual contribution in period 1	.413***	.365***
Average contribution of the remaining group members in period 1	.426***	.532***
Period	.148***	−.169***
cons	3.559***	3.872*
N	10,224	2124
p model	<.001	<.001

Linear mixed effects, choices nested in individuals nested in groups.
Data from periods 2-10.
Hausman test insignificant.
* $p<.05$; ** $p<.01$; *** $p<.001$; + $p<.1$.

Result 2. *The effect of first impressions is not confined to an institution free environment. It does not disappear if those disciplining freeriders must fear revenge*.

4.2. Early intervention

The richer environments provide us with the possibility to test a second implication of broken windows theory. As reported in the introduction, criminal policy has mainly relied on the theory to justify zero tolerance policies. If broken windows are a problem for society, those who break them should be effectively deterred, or so the argument goes. The regressions in Tables 6 and 7 cast doubt on this interpretation of the broken windows metaphor.[10] If punishers must not dread revenge (i.e. in treatment punishment), the main effect of the number of punishment points received in the first period on contributions in later periods is only weakly significant ($p=.081$) once we control for the average contribution of the remaining group members in the first period (model 2). It becomes insignificant if we interact both terms. We then only find a significant positive interaction effect. The more the remaining group members have contributed in the first round, the more effective punishment in that round is in increasing contributions in later rounds (model 3). We resurrect the main effect if we further control for the legitimacy of punishment (model 4). This we do by controlling for the fact that an individual has contributed more than the average of the remaining group members in the first period

[7] $p=.0002$ in model 2, $p=0041$ in model 3.
[8] $p=.0201$.
[9] They are available from the authors upon request.

[10] In the typical design of a public good experiment with punishment, on which we rely both in the reanalysed as in our own data, participants only learn whether and how intensely they have been punished themselves. We can therefore not test for a third interpretation of broken windows theory. We cannot measure the effect of *others* having been punished in the first round.

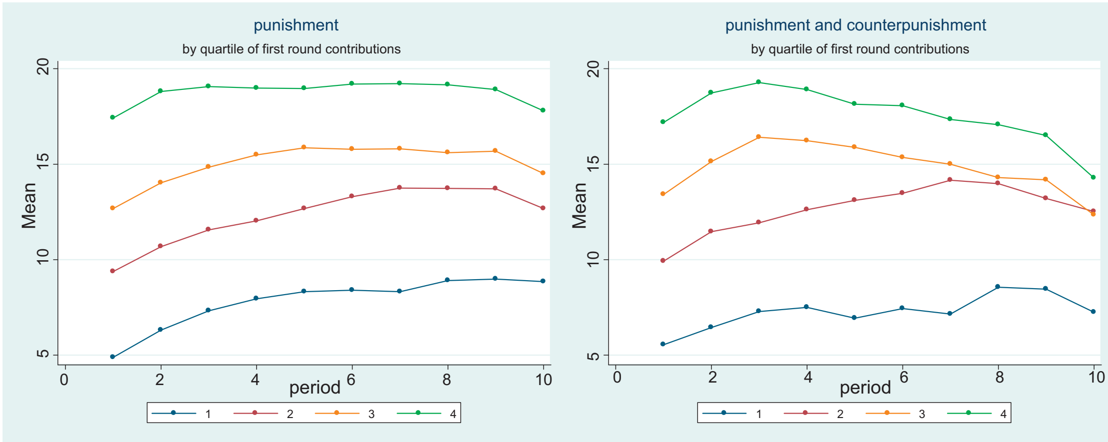

Fig. 5. Effect of first impressions in richer institutional environments.

and, by the interaction term, that this person has been punished nonetheless. Such perverse or antisocial punishment happens in these experiments, and in some locations more often than in others (Herrmann et al., 2008). But even with these additional controls, the effect of early intervention remains small.

One might object that zero tolerance policies put more stress on the certainty rather than on the severity of punishment. In a way, this is reflected in model 6. If we explain contributions in later rounds by the fact that a participant has been punished in the first round, i.e. by a dummy, then we also find a significant effect, at conventional levels, if we control for average contributions of the remaining group members. Yet once we interact both explanatory variables, in this specification neither the main effect of early punishment nor the interaction effect is significant (model 7). Even if we control for the legitimacy of punishment, we only find a weakly significant effect (model 8, $p = .087$).

More strikingly even: once those who have been punished have a chance to strike back, whichever way we specify the regression, we never find a significant effect of early punishment (Table 7).

Comparing coefficients in the models of Table 6, one sees that first impressions have a much stronger effect than early vigilance. Model 5 is easiest to interpret. If a participant has been punished in the first round, however severely, she contributes less than a token more in later rounds. If the remaining group members have only contributed 2 tokens in average in the first round, this already has almost the same effect as any punishment. Fig. 6 adds two more findings. Comparing the error bars, we see that, however long a panel we consider, the effect of early punishment is much more noisy than the effect of first impressions. Moreover while the effect of first impressions virtually stays identical even if we test shorter and shorter panels (confine the sample to choices from the respective period on), the effect of early punishment fades away. From period 7 on, it is no longer statistically different from zero.

We conclude

Result 3. *First impressions are more important than early vigilance. In the long run, the effect of early vigilance fades away, while the effect of first impressions remains stable.*

To be sure, the punishment opportunity is not pointless; comparing Fig. 1 with Fig. 5 one directly sees the strong positive main effect of the punishment threat. But unlike one might have thought, it is not critical whether freeriders have been punished in the first period. The key issue is not early vigilance, but early cooperativeness.

Fig. 7 further illustrates the crucial role of first impressions. Since we have such a rich dataset, we can correlate the average contribution in the first round with the mean contribution in all later rounds.

Table 6
Early intervention vs. first impressions: punishment.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Individual contribution in period 1	.341***	.426***	.433***	.486***	.344***	.433***	.435***	.477***
Amount of punishment received in period 1	.107**	.059+	−.084	.085*				
Punished in period 1					.809***	.583**	.424	.526+
Average contribution of the remaining group members in period 1		.419***	.390***	.359***		.418***	.410***	.373***
avf1*pun1			.012*					
avf1*dumpun1							.016	
Contribution in period 1 above average				−.799*				−.706*
Above*pun1				−.081				
Above*dumpun1								−.044
Period	.148***	.148***	.148***	.148***	.148***	.148***	.148***	.148***
cons	8.905***	3.373***	3.631***	3.727***	8.712***	3.119***	3.182***	3.525***
N	10,224	10,224	10,224	10,224	10,224	10,224	10,224	10,244
p model	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Linear mixed effects, choices nested in individuals nested in groups.
Data from periods 2-10.
avf1: average contribution of remaining group members in period 1.
pun1: amount of punishment received in period 1.
dumpun1: punished in period 1.
above: own contribution in period 1 above average contribution of remaining group members.
Hausman tests insignificant.
* $p < .05$; ** $p < .01$; *** $p < .001$; + $p < .1$.

Table 7
Early vigilance vs. first impressions: punishment and counterpunishment.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Individual contribution in period 1	.224***	.373***	.376***	.415***	.217***	.368***	.373***	.404***
Amount of punishment received in period 1	.065	.053	−.025	.080				
Punished in period 1					.116	.075	−1.258	.052
Average contribution of the remaining group members in period 1		.530***	.519***	.496***		.532***	.475***	.505***
avf1*pun1			.008					
avf1*dumpun1							.113	
Contribution in period 1 above average				−.488				−.539
Above*pun1				−.133				
Above*dumpun1								−.022
Period	−.169***	−.169***	−.169***	−.169***	−.169***	−.169***	−.169***	−.169***
cons	11.737***	3.737*	3.827*	3.870*	11.847***	3.813*	4.448**	3.972*
N	2124	2124	2124	2124	2124	2124	2124	2124
p model	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Linear mixed effects, choices nested in individuals nested in groups.
Data from periods 2-10.
avf1: average contribution of remaining group members in period 1.
pun1: amount of punishment received in period 1.
dumpun1: punished in period 1.
Above: own contribution in period 1 above average contribution of remaining group members.
Hausman tests insignificant.
* $p<.05$; ** $p<.01$; *** $p<.001$; + $p<.1$.

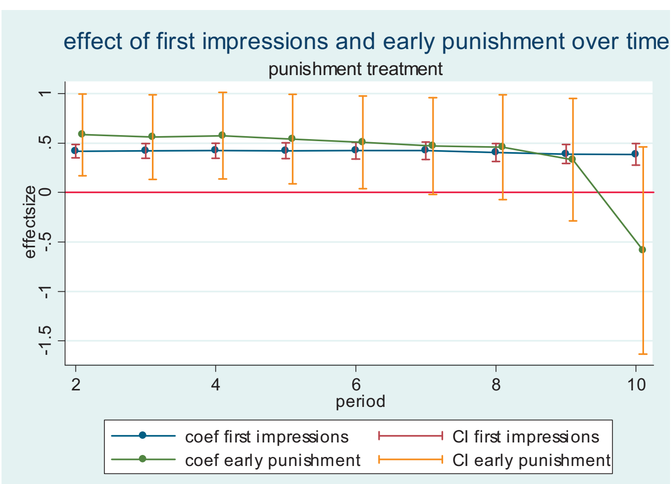

Fig. 6. First impressions and early vigilance in comparison: punishment. Coefficients from model 2 of Table 6 for a series of regressions considering only data from the indicated period on. Model for period 10 only is OLS (since the data is no longer panel data) with standard errors clustered for groups.

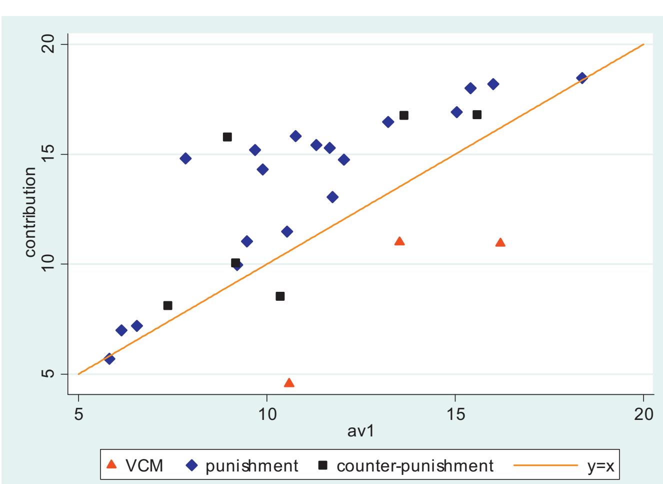

Fig. 7. First impressions determine cooperation.

As one sees, even at this level of aggregation, results seemingly are all over the place. More disturbingly even, it seems that punishment and counterpunishment are pointless. Although in such an environment loyal participants have a chance to discipline freeriders, apparently this does not help them tame antisocial behaviour and improve cooperation. The apparent chaos dissolves once we control for first impressions. All datapoints are in the proximity of the $y=x$ line. If there is no institution, i.e. in the VCM, they are somewhat below this line (red triangles). Otherwise they are usually somewhat above this line. The institution helps participants to even improve over the starting point.

5. Conclusions

Data from 28 experiments conducted all over the world, including five experiments run by us, demonstrate the strong effect of first impressions on cooperation in a linear public good. The average contribution of the remaining group members in the first round determines how much participants contribute to the joint project in later periods. The effect remains discernible until the end game effect kicks in. It does not disappear if one controls for learning. It is present in those who initially contributed less, and in those who contributed more than the average of the other group members. It is present in homogeneous and in heterogeneous groups, i.e. when controlling for the minimum contribution of the remaining group members in the first round. If loyal participants cannot discipline freeriders, despite favourable first impressions, contributions decay over time. If participants are allowed to punish each other, at a cost to themselves, conditional on first impressions contributions stabilize. Their level is determined by first impressions.

Early vigilance, measured by punishment received in the first round, also has a beneficial effect. Yet this effect is much smaller than the effect of first impressions. It only is present if punishers must not fear for revenge. Even absent revenge, the effect of early punishment fades away over time, while the effect of first impressions can even be found in the final period.

The closest real-world analogue to our setting is a person who is new to a neighbourhood. If this person perceives a neat environment, she expects to be treated well if she behaves well herself, and she helps maintain order if she spots signs of erosion. Note that we do not even need normativity to make this prediction. If, in addition, this person is generally willing to abide by the normative expectations prevalent in this community, of course the effect is even stronger. Neither do we need true altruists. All we need is a sufficient proportion of conditional cooperators plus, crucially, the right signals for those who newly enter the community.

In many respects, our experiments have been designed in a way that is congenial to broken windows theory. We observe the minor signs of disorder that this theory posits to be crucial. There are no explicit rules for what "order" means. Normative expectations are idiosyncratic for each context, and have to be inferred from behaviour. In other respects, we put the effect of first impressions to an even harder test: we cannot expect pre-existing social norms to guide behaviour, and there are no public officials who could help the community define expectations, and enforce them if necessary. We deprive participants of any social history, which makes the contributions of others in the first period of interaction a much noisier signal than a decay of order in a previously prosperous neighbourhood. Participants at most loose a bit of experimental money if they spot signs of antisocial behaviour, while they have reason to fear much more in the field. Therefore a vicious cycle should be much more powerful in the field.

Of course, the experimental environment is much poorer and much more artificial than a neighbourhood faced with the onset of disorder or crime. And for sure all we are testing is two components of broken windows theory: the power of first impressions and of early vigilance. Yet these limitations inherent in our method are the price we are paying for the possibility to isolate this effect, and to fully identify it.

With these obvious qualifications, our message to policymakers is straightforward. Money spent on impression management is likely to be money well spent. We can even be more specific. While good first impressions raise overall contributions in the voluntary contribution mechanism, and while they flatten the characteristic negative trend of contributions over time, they are not strong enough to reverse the trend. As many others have shown, both in the lab (Selten, Mitzkewitz, & Uhlich, 1997) and in the field (Ostrom, 1990), for cooperation to be sustainable, vigilance and enforcement are inevitable. However, sanctions alone are also not sufficient. Once we control for first impressions, early vigilance and sanctions at best have a minor beneficial effect. Being determined to prosecute culprits is thus not enough. In a consequentialist perspective, it is more important to manage impressions. Beware of broken windows!

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.irle.2013.07.005.

References

Ashley, R., Ball, S., & Eckel, C. C. (2010). Motives for giving: A reanalysis of two classic public goods experiments. *Southern Economic Journal*, *77*, 15–26.

Blumstein, A. (1995). Youth violence, guns, and the illicit drug industry. *Journal of Criminal Law and Criminology*, *86*, 531–554.

Bowling, B. (1999). The rise and fall of New York murder. Zero tolerance or crack's decline? *British Journal of Criminology*, *86*, 10–36.

Braga, A. A., & Bond, B. J. (2008). Policing crime and disorder hot spots. A randomized controlled trial. *Criminology*, *46*, 577–607.

Braga, A. A., Weisburd, D. L., Waring, E. J., Mazerolle, L. G., Spelman, W., & Gajewski, F. (1999). Problem-oriented policing in violent crime places. A randomized controlled experiment. *Criminology*, *37*, 541–580.

Brekke, K. A., Hauge, K. E., Lind, J. T., & Nyborg, K. (2009). *Playing with the Good Guys: A Public Good Game with Endogenous Group Formation*. http://folk.uio.no/karineny/files/GoodGuys.pdf

Cinyabuguma, M., Page, T., & Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, *89*, 1421–1435.

Corman, H., & Mocan, N. (2005). Carrots, sticks, and broken windows. *Journal of Law and Economics*, *48*, 235–266.

Croson, R. T. A., Fatas, E., & Neugebauer, T. (2005). Reciprocity, matching and conditional cooperation in two public goods games. *Economics Letters*, *87*, 95–101.

Croson, R. T. A., Fatas, E., & Neugebauer, T. (2008). *The Effect of Excludability on Team Production*. http://www.economics.hawaii.edu/research/seminars/08-09/11_07_08b.pdf

Croson, R. T. A., & Shang, J. (2008). The impact of downward social information on contribution decisions. *Experimental Economics*, *11*, 221–233.

Cruz Melendez, M. (2006). Moving to opportunity & mending broken windows. *Journal of Legislation*, *32*, 238–262.

Denant-Boèment, L., Masclet, D., & Noussair, C. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, *33*, 145–167.

Fagan, J. (2008). Crime and neighborhood change. In R. Rosenfeld, & A. S. Goldberger (Eds.), *Understanding crime trends* (pp. 81–126). Washington: National Academy Press.

Farrington, D. P. (2003). A short history of randomized experiments in criminology. A meager feast. *Evaluation Review*, *27*, 218–227.

Farrington, D. P. (2006). Key longitudinal-experimental studies in criminology. *Journal of Experimental Criminology*, *2*, 121–141.

Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology. What have we learned in the last two decades? *Journal of Experimental Criminology*, *1*, 9–38.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*, 980–994.

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public good experiments. *American Economic Review*, *100*, 541–556.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*, 397–404.

Frey, B., & Meier, S. (2004). Social comparisons and pro-social behavior: Testing conditional cooperation in a field experiment. *American Economic Review*, *94*, 1717–1722.

Funk, P., & Kugler, P. (2003). Dynamic interaction between crimes. *Economics Letters*, *79*, 291–298.

Gächter, S., & Thöni, C. (2007). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, *3*, 303–314.

Geller, A. (2007). *Neighborhood disorder and crime. An analysis of broken windows in New York City*. http://ssrn.com/abstract=1079879

Gunnthorsdotir, A., Houser, D., & McCabe, K. (2007). Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior & Organization*, *62*, 304–315.

Harcourt, B. (1998). Reflecting on the subject. A critique of the social influence conception of deterrence, tne broken windows theory, and order-maintenance policing New York style. *Michigan Law Review*, *97*, 291–389.

Harcourt, B. (2001). *Illusions of order. The false promise of broken windows policing*. Boston: Harvard University Press.

Harcourt, B. (2005). Policing L.A.'s skid row: Crime and real estate development in downtown Los Angeles. An experiment in real time. *University of Chicago Legal Forum*, *2005*, 325–404.

Harcourt, B., & Ludwig, J. (2006). Broken windows: New evidence from New York City and a five-city social experiment. *University of Chicago Law Review*, *73*, 271–320.

Heckathorn, D. D. (1989). Collective action and the second-order free-rider problem. *Rationality and Society*, *1*, 78–100.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*, 1362–1367.

Karmen, A. (2000). *New York murder mystery. The true story behind the crime crash of the 1990*. New York: New York University Press.

Katz, C. M., Webb, V. J., & Schaefer, D. R. (2001). An assessment of the impact of quality-of-life policing on crime and disorder. *Justice Quarterly*, *18*, 825–876.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science*, *322*, 1681–1685. http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=35903192&site=ehost-live

Kelling, G. L., & Coles, C. M. (1996). *Fixing broken windows. Restoring order and reducing crime in our communities*. New York: Martin Kessler Books.

Kelling, G. L., & Sousa, W. H. (2001). *Do police matter? An analysis of the impact of New York City's police reforms*. New York: Manhattan Institute.

Keser, C., & van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, *102*, 23–39.

Lochner, L. (2007). Individual perceptions of the criminal justice system. *American Economic Review*, *97*, 444–460.

Messner, S. F., Galea, S., Tardiff, K., Tracy, M., Bucciarelli, A., Piper, T. M., et al. (2007). Policing, drugs, and the homicide decline in New York City in the 1990. *Criminology*, *45*, 385–414.

Nikiforakis, N. S. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*, 91–112.

Novak, K. J., Hartman, J. L., Holsinger, A. M., & Turner, M. G. (1999). The effects of aggressive policing of disorder on serious crime. *Policing: An International Journal of Police Strategies & Management*, *22*(2), 171–194.

Ostrom, E. (1990). *Governing the commons. The evolution of institutions for collective action*. Cambridge, NY: Cambridge University Press.

Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public goods experiments. Reciprocity, mimicry and efficiency. *Economic Journal*, *115*, 1032–1053.

Rosenfeld, R., Fornango, R., & Rengifo, A. (2007). The impact of order-maintenance policing on New York City homicide and robbery rates. 1988–2001. *Criminology*, *45*, 355–384.

Sampson, R. J., & Cohen, J. (1988). Deterrent effects of the police on crime. A replication and theoretical extension. *Law and Society Review*, *22*, 163–190.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing 'Neighborhood Effects'. Social processes and new directions in research. *Annual Review of Sociology*, *28*, 443–478.

Sampson, R. J., & Raudenbush, S. W. (1999). Systematic social observation of public spaces. A new look at disorder in urban neighbourhoods. *American Journal of Sociology*, *105*, 603–651.

Schuerman, L., & Kobrin, S. (1986). Community careers in crime. *Crime and Justice*, *8*, 67–100.

Selten, R., Mitzkewitz, M., & Uhlich, G. R. (1997). Duopoly strategies programmed by experienced players. *Econometrica*, *65*, 517–555.

Skogan, W. G. (1990). *Disorder and decline. Crime and the spiral of decay in American neighborhoods*. New York, Toronto: Free Press; Colleir Macmillan Canada.

Sousa, W. H., & Kelling, G. L. (2006). Of "Broken Windows", criminology, and criminal justice. In D. L. Weisburd, & A. A. Braga (Eds.), *Police innovation. Contrasting perspectives*. New York: Cambridge University Press.

Taub, R. P., Taylor, D. G., & Dunham, J. D. (1984). *Paths of neighborhood change. Race and crime in Urban America*. Chicago, IL: University of Chicago Press.

Taylor, R. B. (2001). *Breaking away from broken windows. Baltimore neighborhoods and the nationwide fight against crime, grime, fear, and decline*. Boulder, Colo: Westview Press.

Telep, C. W. (2009). Citation analysis of randomized experiments in criminology and criminal justice. A research note. *Journal of Experimental Criminology*, *5*, 441–463.

Wagers, M. L. (2008). Broken windows policing. The LAPD experience. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *68*(8-A), 3603.

Wilson, J. Q., & Boland, B. (1978). The effect of the police on crime. *Law and Society Review*, *12*, 367–390.

Wilson, J. Q., & Kelling, G. L. (1982). Police and neighborhood safety. Broken windows. *Atlantic Monthly*, *127*, 29–38.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, *51*, 110–116.

Zimbardo, P. (1969). The human choice. Individuation, reason, and order versus deindividuation, impulse, and chaos. *Nebraska Symposium on Motivation*, *17*, 237–307.

Zimring, F. E. (2007). *The great American crime decline*. Oxford, NY: Oxford University Press.