

Revisiting Broken Windows: The Role of Neighborhood and Individual Characteristics in Reaction to Disorder Cues

Beate Volker

University of Amsterdam

Abstract: The influential "broken windows" theory proposes that disorder cues in neighborhoods trigger littering and other antisocial behavior. Until now, the theory has been empirically tested only on a small scale and restricted to just one specific area. In this study, I investigated the effect of disorder cues on individual behavior once more, replicating and extending the original field experiments by Keizer, Lindenberg, and Steg (2008 and 2013). The data from 12,528 individuals were collected in 84 field experiments conducted in 33 neighborhoods. The results, based on multilevel techniques for binary data, show that the absolute effect of cues is smaller than originally thought and that neighborhood and individual characteristics moderate cue effects.

Keywords: broken windows theory; field experiments; cue effects; context effects

SINCE the first publication (Wilson and Kelling 1982) on the subject, the "broken windows" theory has been influential in explaining neighborhood crime and deterioration. The theory assumes that minor signs of physical disorder, such as a broken window not being fixed, attract not only more physical disorders but also serious forms of social disorders, such as street violence, robbery, or even violent offenses. Importantly, the effects of disorder cues are assumed to be independent of contextual variation. For example, it is alleged that a disorder cue affects individual behavior regardless of the neighborhood in which the cue is presented: "...if a window in a building is broken and is left unrepaired, all the rest of the windows will soon be broken. This is as true in nice neighborhoods as in run-down ones" (Wilson and Kelling 1982:2). However, the claim that the effect of disorder cues is independent of contextual characteristics is empirically and theoretically implausible. In his famous but never scientifically published field experiment from the 1960s, Zimbardo showed that neighborhood status and composition have a clear effect on people's reactions to cues¹ (e.g., Chu 2007; Sampson 2012a). Furthermore, both the theory of social disorganization and the large body of socioecological literature on context effects in general have argued (and have convincingly shown) that social contexts—such as neighborhoods—affect many individual outcomes and behaviors (for an overview, see Sampson, Morenoff, and Gannon-Rowley 2002). This article investigates the effects of contextual characteristics on disorder cues while considering individual differences.

Citation: Volker, Beate. 2017. "Revisiting Broken Windows: The Role of Neighborhood and Individual Characteristics in Reaction to Disorder Cues" *Sociological Science* 4: 528-551.

Received: July 22, 2017

Accepted: August 8, 2017

Published: October 11, 2017

Editor(s): Jesper Sørensen, Mario Small

DOI: 10.15195/v4.a22

Copyright: © 2017 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

Does Physical Disorder Prompt Social Disorder?

The proposed association between physical and social order in neighborhoods has been extensively discussed in the social sciences (e.g., Harcourt 2006) and is considered established (albeit serious) criticism. Calls for policy measures and large-scale policy implementations as well as popular arguments have been based on the theoretical assumption supporting such an association (see, among others, Kelling and Coles [1996] and Gladwell [2000] on tipping points and the effects of small actions on succeeding events). However, systematic tests of the association have been scarce. The work of Sampson and Raudenbush (1999), who found a modest correlation between physical and social disorder, is one of the first systematic inquiries. Other studies have concluded that the evidence for the effects of broken windows is mixed and does not allow for definite conclusions on causality (Harcourt 2006).

More recently, Keizer et al. (2008 and 2013) tested the theory more thoroughly. They argued that the evidence for the broken windows theory's claims is merely correlational. Indeed, until this work, studies had failed to inquire about the link between physical cues and social order at an individual level. The authors set up a series of field experiments that tested whether disorder cues affected norm-conforming behavior. For example, they counted how often people deposited a letter that had been dropped next to a mailbox depending on whether the immediate environment was clean or littered. Another experiment involved counting whether a fallen bike was picked up in the presence or absence of disorder cues. In total, the researchers conducted eight different experiments—five in 2008 and three in 2013—in the city of Groningen.² The results showed overwhelming confirmation of the effect of small cues on individual behavior. The results are depicted in Table A1 in the online supplement.

Despite the creativity of the field experiments, their design, analyses, and conclusions have been criticized (Wicherts and Bakker 2014): the findings' reliability and validity, *inter alia*, were questioned because the field experiments did not control for individual self-selection into different experimental conditions. For example, age and gender are individual characteristics that might correlate with littering behavior and with the perception of litter (see Sampson and Raudenbush [2004], who studied individual differences in perceiving disorder). In addition, the sample size was small, the analyses were based on the assumption that individuals acted independently of each other, and no corrections for multiple testing were applied. In particular, the latter led to type I error inflation. In addition, the experiments have been tested separately, and an integrative analysis of the data has not been conducted. Consequently, an integrative meta-analysis conducted by Wicherts and Bakker (2014:402) cast doubt on the results, and the authors concluded that the effect of cues "remains to be seen" because the analyses so far did not involve adequate methodological controls for confounding factors and observer biases.³

This Study

This study builds upon this discussion. I deliberately inquired into socially desired behavior as small, noncostly acts of kindness or prosocial behavior, thereby building upon seminal studies on the effects of area on littering (Reiter and Samuel 1980; Cialdini, Reno, and Kallgren 1990; De Kort, McCalley, and Midden 2008; Schultz et al. 2011; Weaver 2015). The article contributes to the existing literature in three respects. First, I investigated the effect of disorder cues on individual behavior in a large-scale, field experiment setting. Second, the shortcomings of a field experiment (Shadish, Cook, and Campbell, 2002)—that is, an experiment in which individuals are not randomly assigned to treatment conditions—are compensated to the maximum extent possible by collecting information on individual characteristics and controlling for these characteristics in multivariate analyses. Third, it is studied whether the neighborhood contextual and compositional characteristics, together with the presence or absence of disorder cues, influence individual behavior. This way, the study provides a new and comprehensive test of the assumption of the broken windows theory on the general effect of environmental cues for individual behavior.

Expectations: Cues, Neighborhoods, and Individuals

Following Keizer et al. (2008), I argue that disorder cues stimulate norm-violating behavior, and order cues stimulate behavior that is in agreement with general social norms. It is relatively well established that social behavior depends on subtle cues (e.g., Haley and Fessler 2005). Less clear, however, is how the wider context in which a cue is presented affects individual reactions, and Keizer et al. (2008 and 2013) did not inquire into the moderation of cue effects by social contexts. Furthermore, reaction to a cue arguably depends on the specific characteristics of an individual (see below).

Concerning the influence of the neighborhood context, the degree to which a neighborhood is well functioning and advantaged can influence cue effects. It has been shown that being exposed to cues leads to familiarity (see Zajonc 1968 and Bornstein and Craver-Lemley 2004). If one is familiar with disorder in the environment, an additional cue bears no new information and cannot be seen as a trigger for behavior (see Keuschnigg and Wolbring 2015). Additionally, but not completely in line with this argument, it has been shown that neighborhood composition predicts perception of neighborhood disorder and other problems (Sampson and Raudenbush 2004; Sampson 2012a). In particular, the perception of disorder is enhanced in neighborhoods with many minority groups and a relatively high poverty rate. Notwithstanding these compositional effects on perception of (dis)order, a disorder cue presented in a neighborhood that is clearly disorganized probably does not attract attention and trigger behavior; it might not even be noticed. However, when presented in a well-functioning and well-organized neighborhood, a cue is more noticeable, and it can be a signal that influences behavior. Yet in such well-organized neighborhoods, disorder cues might not have the effect that one would expect, according to the theory of broken windows. This theory argues that

disorder cues stimulate norm-violating behavior because they signal that “nobody cares,” implying that behavior will not be sanctioned. Instead in well-organized neighborhoods, cue effects might go beyond the “nobody cares” impression. Rather, they might trigger behavior that reinstalls order: that is, intervening on behalf of the collective good and behaving prosocially *although* there are signs of disorder. Norms about social order might in general be more pronounced in well-functioning and advantaged neighborhoods, whereas there is little room for norms such as, “We care for what happens in our neighborhood” in rundown, deteriorated areas. Hence, in such advantaged neighborhoods, disorder cues might stimulate (rather than inhibit) “cleaning” behaviors.

Volker et al. (2015) demonstrated that (as they call it) other-regarding behavior, such as posting a letter that seems to be lost, depends on the degree of collective efficacy in a neighborhood, among other things. Collective efficacy can be understood as the shared norm on behalf of which people will intervene for the collective good. Likewise, Koopmans and Veit (2014) showed decreased cooperative behavior in ethnically heterogeneous neighborhoods. In a more advantaged neighborhood, other-regarding behavior is higher and people might be more aware of the environment in general (Sampson 2012b; Volker et al. 2015). In short, it is expected that in more advantaged neighborhoods, people will show more prosocial behavior (independent of cues), and in addition, disorder cues can have reverse effects in these neighborhoods.

Concerning specific individual characteristics, theories of neighborhood effects such as the social disorganization theory (Shaw and MacKay 1969) as well as most field experiments that have been undertaken to establish neighborhood effects leave open important questions: What individual characteristics explain the behavior that causes the neighborhood effect in question? What type of individual behavior is responsible for neighborhood cohesion? How does a potential offender notice the level of collective efficacy in a neighborhood? In other words, the link between microbehavior and the macroeffect in neighborhood research is often not established. Moreover, it is plausible to expect individual heterogeneity in the reaction to cues. For example, women might be socialized to behave more prosocially in general. Also, the elderly might conform more to norms of prosocial behavior. In addition, when presented with a disorder cue, women as well as the elderly might feel inclined to reinstall order. Finally, younger people might litter more easily and feel more easily stimulated by disorder cues to show antisocial behavior. Consequently, independent of disorder cues, women and the elderly are expected to show more prosocial behavior, and they might even be triggered to behave prosocially when presented with such a cue, whereas younger individuals might react to such a cue by behaving antisocially.

In summary, in this article, the expectations concerning the general effects of disorder cues on prosocial behavior are broadened in the following respects. First, it is argued that the wider context of the neighborhood impacts the degree to which a cue influences behavior. People who are already more exposed to disorder in their daily life will not react to such a cue. For people who are usually not exposed to disorder in the setting in which they live, a disorder cue might enhance rather than suppress norm-conforming behavior: they might feel inclined to reinstall

physical or social order rather than violate more norms of social order. Second, it is expected that people in more advantaged neighborhoods behave in a more norm-conforming way (i.e., prosocially) in general than in less advantaged neighborhoods, independent of the cue. Third, this also holds for specific individual categories: because of their social roles and activities, some people, particularly women and the elderly, will show more prosocial behavior in general; furthermore, disorder cues might also trigger behavior toward the reinstallation of social order than norm violation in these groups. In short, and more technically, next to the main effects of cues, I expect main effects of specific neighborhood characteristics and individual categories as well as moderated cue effects by neighborhoods' and individuals' characteristics.

Data

In total, 84 field experiments have been conducted in 33 neighborhoods in five cities in the Netherlands: Amsterdam, Amersfoort, Utrecht, Rotterdam, and The Hague. Neighborhoods were chosen randomly from the sample of residential neighborhoods in and around the center of these cities. Additionally, the original letter experiment of Keizer et al. (2008) has been replicated in Groningen on the same street as the one used in the original experiment.

Each experimental condition was applied to approximately 70 passersby whose actions were observed and whose characteristics were documented by three to four observers. A sample observer sheet and sample photographs of the research site are provided in the supplementary materials (Appendices 2 and 3). If the observers were uncertain about, for example, the age of a person, they were instructed to both report the difference and to discuss their best guesses. In addition, the observers were asked to report whether a person had not seen the cue.

Observers were instructed to take a position approximately 10 meters from the mailbox and, holding a mobile phone, make it appear that they were busy and not watching the area. In this manner, the data from 12,528 individuals were collected.

Neighborhood characteristics were provided by Statistics Netherlands (CBS) (2016).

Experimental Setup

For practical reasons, this study focused on three experiments: the letter experiment, the bike experiment, and the money-in-letter experiment (see Keizer et al. 2008 and 2013).

- Letter experiment: A stamped letter with an address and sender information was dropped on the sidewalk close to a public mailbox. In the order condition, it was ensured that the environment around the mailbox was clean and showed no signs of litter. In the disorder condition, two or three garbage bags were stored close to the mailbox, with Coca-Cola cans and orange peels visible.

- Bike experiment: A bike was put on the sidewalk so it would look like it had fallen. Conditions were the same as in the letter experiment.
- Money experiment: Conditions were the same as in the letter experiment, but the envelope had a window, and a 5-Euro bill was visible.

All experiments were conducted in January, February, and March 2015 on weekdays either in the morning or at noon, accounting for light and weather conditions. If it was rainy or windy, the experiment was postponed. In addition, the experiments had to stop at 4:30 p.m. because of worsening light conditions.

The Research Sites: Neighborhoods in Dutch Cities

The 33 neighborhoods in the Netherlands were part of five cities.. All neighborhoods were purely residential and close to the city center (not more than 3 km). Four out of the five cities are the biggest in the country (Amsterdam, Rotterdam, Den Haag, and Utrecht), and they are all located in the main part, the Randstad. The number of inhabitants in these cities varies from 340,000 in Utrecht to 840,000 in Amsterdam (Statistics Netherlands [CBS] 2016; rounded figures); thus, the residential neighborhoods in the city centers are densely populated. In contrast, the fifth city, Amersfoort, is considerably smaller (154,000 inhabitants) and is located in the southeast of the country, close to the border with Germany, implying that it is out of the "main scene."

The cities vary according to important characteristics that predict social cohesion and collective efficacy (e.g., Sampson 2010). For example, residential turnover is highest in the cities of Amsterdam and Utrecht, partially because of the high number of students who live there. At the same time, the proportion of married people is lowest in Amsterdam and highest in Amersfoort, where the proportion of the elderly is much larger compared with the other cities in the study. Furthermore, housing prices are highest in Amersfoort and lowest in Rotterdam. Regarding the number of foreigners, the percentage of non-Western foreigners is highest in Rotterdam and The Hague whereas the percentage of Western foreigners is highest in Amsterdam and Utrecht. Within the cities, neighborhoods also clearly differ in these regards. In general, differences between neighborhoods are sometimes even larger than differences between cities.

In addition to studying neighborhoods in different cities, I also replicated the original field experiment by Keizer et al. (2008) on the very same street in Groningen. Groningen is located in the northern part of the country, with 200,500 inhabitants, which is comparable with Amersfoort; like the latter, Groningen is located outside the urban concentration area, the Randstad.

Measurements and Analytical Strategy

Dependent Variables

The dependent variables were small acts of kindness, prosocial behavior, or other-regarding behavior (Volker et al. 2015), the same as in the experiments of Keizer et al. (2008 and 2013). Specifically, they measured the following:

- (in the letter experiment) whether a passenger posts a letter into the mailbox;
- (in the money experiment) whether a passenger deposits the letter without taking the money; and
- (in the bike experiment) whether passengers pick up a fallen bike.

These other-regarding actions were coded with 1 if they were undertaken (as opposed to 0 if nothing was done). Both "other-regarding behavior" and "prosocial behavior" are used to indicate these actions.

Independent Variables

For each person, it was coded whether he or she was in (1) the disorder or (2) the order condition of the experiments. In the disorder condition, the immediate environment around the letter was littered whereas it was clean in the order condition.

Individual-level control variables. From the individual passengers, the observers collected the following information:

- gender (0 = male; 1 = female);
- age (1 = up to 25 years; 2 = 26 to 35 years; 3 = 36 to 45 years; 4 = 46 to 55 years; 5 = 56 to 65 years; 6 = older than 65 years);
- appearance of having a migration background (0 = no; 1 = yes); and
- individual passenger (0) or group (1).

In the analyses, the gender of people and group or individual status were combined into woman alone, men alone, group of men, group of women, and gender-mixed group.

Neighborhood composition and contextual control variables. I controlled for neighborhood composition in terms of ethnicity, age, and family status.

- Ethnic composition was measured as the percentage of Western and non-Western foreigners in a neighborhood. Statistics Netherlands considers inhabitants with at least one foreign-born parent to be foreigners.
- Age composition was measured by six dummy variables indicating the percentage of residents who are (1) below 14 years of age, (2) 15 to 25 years of age, (3) 26 to 45 years of age, (4) 46 to 65 years of age, and (5) above 65 years of age.

- Family status was measured by the percentage of married couples in a neighborhood.

In addition, I controlled for neighborhood contextual characteristics. Following Shaw and McKay (1969), I considered high ethnic heterogeneity, low price of houses or high percentage of houses for rent, and high residential fluctuation as proxies for neighborhood disorganization and low collective efficacy (see also Sampson 2010).

- Relative residential fluctuation was measured by Statistics Netherlands as the sum of movers within a region plus the sum of those who left and those who came to the region divided by 2.
- The percentages of houses that are owned compared to rented properties were also included in the analyses to indicate neighborhood advantages.
- Finally, neighborhood status scores provided by the Netherlands Institute for Social Research were included in the analyses. These scores measure relative neighborhood advantage in terms of income, education, and position in the labor market (see Knol, Boelhouwer, and Veldheer 2012).

All neighborhood indicators are provided by Statistics Netherlands and are publicly available. Appendix 4 in the online supplement provides descriptive information on the key individual and neighborhood variables.

Neighborhood delineation. Neighborhoods were delineated by postal codes. In the Netherlands, postal codes consist of up to six digits. On average, six-digit areas consist of approximately 40 addresses, five-digit areas consist of approximately 400 addresses, and four-digit areas consist of 4,000 addresses. For this article, information on five-digit areas is used, which was combined with the information provided by Statistics Netherlands.

Analytical strategy. First, the analysis has been replicated by using statistical techniques like those applied by Keizer et al. Next, it accounted for both the nested structure of the data and the nature and distribution of the dependent variable. Multilevel fixed-effect models for binary data have been estimated.⁴ After estimating the empty model, the order and disorder variable was added, followed by the individual variables, neighborhood variables, and the interaction terms to assess the moderated cue effects. To show that cue effects differ across neighborhoods and individuals, models for two different cues (disorder and order) were also established.

Two sets of variables were entered in the models on the effects of neighborhood characteristics. First, I followed Shaw and MacKay (1969) and estimated models on the influence of neighborhood disorganization, including the variables of residential fluctuation, ethnic heterogeneity (the percentage of non-Western migrants), and neighborhood wealth (owned property). Second, compositional characteristics were included, such as the number of married people, the number of elderly people, and the status of the neighborhood in terms of labor market position and the income of its residents.

Because the number of neighborhoods was limited, the models on the effect of neighborhood conditions and their interactions with the disorder cues were estimated in separate analyses (see Snijders 2005). The parameters' significance was established by using Wald tests. The data were analyzed with Stata 15, and they are available upon request. Furthermore, the data on the different experiments were pooled for the analyses (see below), and the analyses for the different types of experiments are also available upon request.

The binomial model controlling for cue, type of experiment, and individual characteristics was based on the following equation:

$$\text{other regarding behavior}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\begin{aligned} \text{logit}(\pi_{ij}) = & \beta_0 \text{cons} + \beta_1 \text{ordercue}_{ij} + \beta_2 \text{woman_alone}_{ij} \\ & + \beta_3 \text{gendermixed_group}_{ij} + \beta_4 \text{group_men}_{ij} + \beta_5 \text{group_women} \\ & + \beta_6 \text{age26} - 35_{ij} + \beta_7 \text{age36} - 45_{ij} + \beta_8 \text{age46} - 55_{ij} \\ & + \beta_9 \text{age56} - 65_{ij} + \beta_{10} \text{age65+}_{ij} + \beta_{11} \text{foreignappearance}_{ij} \end{aligned}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [\sigma_{u0}^2]$$

$$\text{var}(\text{other regarding behavior}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

The intraclass coefficient was established by

$$ICC = \frac{\sigma^2}{\sigma^2 + 3.29}$$

where σ^2 is the variance at the neighborhood level (Snijders and Bosker [1999] 2011; code "estat ICC" in Stata 15.).

Results

In general, other-regarding behavior occurred in 4.7 percent ($n = 594$) of all cases, 5.9 percent ($n = 374$) of the cases in the order condition ($n = 6248$), and 3.5 percent ($n = 220$) of the cases in the disorder condition ($n = 6280$). This difference is significant: $\chi^2(1/12528) = 42.743, p < 0.001$. At the neighborhood level, a paired t test revealed significant mean differences in other-regarding behavior between the order and the disorder condition ($t = 2.899$, degrees of freedom = 32, $p < 0.006$). Although the absolute numbers are small, this indicates a clear, relative treatment effect: in the order condition, other-regarding behavior was about 40 percent higher compared with the disorder condition. Cohen's d , indicating the strength of the effect, was 0.12, which is a very small effect. Note that the treatment effect is much lower compared with the effect found in the field experiments by Keizer and colleagues (see Appendix 1 in the online supplement). Table A1 shows that the order cue

Table 1: Other-regarding behavior by cue and sort experiment.

Postal code	Number of experiment	Letter experiment		χ^2	p value	Odds ratio	N
		Other-regarding behaviour under condition of: Disorder % (n)	Order % (n)				
1012G	1	2.9 (2)	0.0 (0)	2.029	0.154	0.493	140
	2	0.0 (0)	0.0 (0)	-			140
1012J	3	0.0 (0)	0.0 (0)	-			144
	4	0.0 (0)	0.0 (0)	-			138
1015M	5	7.1 (5)	0.0 (0)	5.185	0.023	0.481	135
	6	5.7 (4)	0.0 (0)	4.118	0.042		0.485
1061A	7	1.4 (1)	12.9 (9)	6.892	0.009	10.180	130
	8	11.8 (8)	5.7 (4)	1.590	0.207		0.455
1092A	9	0.0 (0)	0.0 (0)	-			180
	10	0.0 (0)	3.6 (2)	1.853			0.173
2512G	11	1.1 (1)	2.2 (2)	0.339	0.560	2.023	177
	12	1.1 (1)	1.1 (1)	-			-
2515L	13	0.0 (0)	0.0 (0)	-			180
	14	0.0 (0)	0.0 (0)	-			-
2516K	15	0.0 (0)	0.0 (0)	-			179
	16	1.5 (1)	0.0 (0)	1.054			0.305
2525E	17	0.0 (0)	4.3 (3)	3.066	0.080	0.489	137
	18	0.0 (0)	11.1 (7)	7.421			0.006
2526G	19	0.0 (0)	0.0 (0)	-			180
	20	0.0 (0)	0.0 (0)	-			-
2526H	21	0.0 (0)	0.0 (0)	-			140
	22	0.0 (0)	0.0 (0)	-			-
2526J	23	0.0 (0)	0.0 (0)	-			140
	24	0.0 (0)	0.0 (0)	-			-
2533J	25	7.1 (5)	7.2 (5)	0.001	0.981	1.016	129
	66	11.4 (8)	17.1 (12)	0.933			0.334
2562A	26	2.9 (2)	0.0 (0)	2.029	0.154		138
	27	0.0 (0)	0.0 (0)	-			-
2564E	28	11.4 (8)	4.3 (3)	2.467	0.116	0.347	129
	59	12.9 (9)	18.8 (15)	0.965			0.326
3021B	29	0.0 (0)	0.0 (0)	-			156
	30	0.0 (0)	1.4 (1)	1.021			0.312
3535T	31	3.8 (3)	1.3 (1)	0.949	0.330	0.338	155
	32	3.8 (3)	1.4 (1)	0.820			0.365
3074D	33	1.4 (1)	2.9 (2)	0.341	0.559	2.029	140
	34	2.9 (2)	2.9 (2)	-			-
3511X	35	0.0 (0)	1.1 (1)	0.961	0.327	0.509	176
	36	4.7 (7)	0.0 (0)	5.039			0.025
3515L	37	0.0 (0)	2.9 (2)	2.029	0.154	0.493	140
	38	1.4 (1)	2.8 (2)	0.326			0.568
3527K	39	0.0 (0)	15.7 (11)	11.938	0.001	0.457	140
	40	5.7 (4)	21.4 (15)	7.368			0.007
3551D	41	0.0(0)	0.0 (0)	-			180
	42	4.3 (3)	2.9 (2)	0.207			0.649

Table 1 continued

Postal code	Number of experiment	Letter experiment		χ^2	<i>p</i> value	Odds ratio	<i>N</i>
		Other-regarding behaviour under condition of: Disorder % (<i>n</i>)	Order % (<i>n</i>)				
3562K	43	2.9 (2)	18.6 (13)	9.035	0.003	7.754	140
	44	2.9 (2)	15.7 (11)	6.869	0.009	6.339	140
3571L	45	13.7 (10)	16.9 (13)	0.293	0.588	1.280	150
	46	12.9 (9)	14.1 (9)	0.042	0.838	1.109	134
3812D	47	4.4 (4)	2.2 (2)	0.712	0.399	0.483	183
	48	5.7 (4)	1.4 (1)	1.867	0.172	0.239	140
3813D	49	12.9 (9)	40.0 (28)	13.262	0.000	4.519	140
	50	14.5 (10)	40.0 (28)	11.381	0.001	3.933	139
3813P	51	1.4 (1)	4.2 (3)	1.000	0.317	3.044	181
	52	1.4 (1)	4.3 (3)	1.059	0.304	3.130	140
3815G	53	8.6 (6)	22.9 (16)	5.231	0.022	3.111	139
	54	8.6 (6)	21.4 (15)	4.538	0.033	2.909	140
3818E	55	7.8 (7)	2.2 (2)	2.982	0.084	0.266	181
	56	10.4 (7)	1.1 (1)	4.773	0.029	0.130	140
3823E	57	11.4 (8)	28.6 (20)	6.429	0.011	3.100	140
	58	11.1 (9)	38.6 (27)	15.594	0.000	5.023	151
2564G	60	2.0 (1)	16.0 (8)	5.983	0.014	9.333	100
	61	10.0 (7)	16.3 (13)	1.262	0.261	1.746	150
2564V	62	7.5 (3)	2.5 (1)	0.913	0.344	0.316	136
	63	8.6 (6)	12.9 (9)	0.672	0.412	1.574	140
2591V	64	1.1 (1)	5.0 (3)	1.069	0.301	3.158	121
	65	1.3 (1)	13.3 (8)	7.106	0.008	10.615	130
Bike and money experiments							
1092A	67 (B)	0.0 (0)	1.1 (1)	0.983	0.321	0.582	178
	68 (B)	0.0 (0)	3.6 (2)	1.852	0.173	0.621	105
2512G	69 (B)	1.1 (1)	4.4 (4)	1.851	0.174	4.140	180
	70 (B)	1.1 (1)	1.1 (1)	-	-	-	180
2515L	71 (B)	0.0 (0)	0.0 (0)	-	-	-	180
2516K	72 (B)	0.0 (0)	0.0 (0)	-	-	-	179
	73 (B)	1.5 (1)	0.0 (0)	1.054	0.305	0.485	133
2526G	74 (B)	0.0 (0)	0.0 (0)	-	-	-	193
	75 (B)	0.0 (0)	1.6 (1)	0.942	0.332	0.513	118
2562A	76 (B)	2.9 (2)	0.0 (0)	2.029	0.154	0.493	140
3818E	77 (B)	7.8 (7)	3.3 (3)	1.741	0.187	0.404	181
3021B	78 (B)	0.0 (0)	0.0 (0)	-	-	-	156
3812D	79 (B)	1.1 (1)	2.2 (2)	0.328	0.567	2.000	183
3511X	80 (B)	0.0 (0)	0.5 (1)	0.955	0.328	1.010	195
3551D	81 (B)	0.0 (0)	0.0 (0)	-	-	-	180
3571L	82 (M)	0.0 (0)	0.0 (0)	-	-	-	150
3035T	83 (M)	13.7 (10)	22.1 (17)	0.169	0.681	1.383	155
2591V	84 (M)	3.8 (3)	2.6 (2)	1.069	0.301	3.158	121

Note: B = bike experiment, M = money experiment; see text for explanation.

enhanced prosocial behavior by at least 100 percent; sometimes, a more than 600-percent increase was observed.

Table 1 presents the percentage of other-regarding behavior per each experimental condition and neighborhood and the related χ^2 statistic. Significant cue effects have been found in 17 out of the 84 experiments (p values are printed in bold). However, in three experiments, other-regarding behavior was significantly higher in the disorder condition. Furthermore, on 22 occasions, other-regarding behavior was observed slightly more in the order condition, but the effect was not significant. In 25 experiments, none or the same amount of other-regarding behavior was shown, regardless of the experimental condition.

Considerable differences emerged among the five cities, as shown in Figure 1. Other-regarding behavior was highest in the city of Amersfoort and lowest in Rotterdam. In Amsterdam, other-regarding behavior was shown most often in the disorder condition. Because of these differences, models without the outliers have also been estimated, but this did not change the main conclusions.

Tables 2 and 3 present the binomial, multilevel, fixed-effect regression models. $-2LL$ for the empty model is -2015.64 . Table 2 shows the cue effect (model 1 [M1]) controlled for type of experiment. The overall cue effect has an odds ratio of 1.8, indicating that the odds that other-regarding behavior is shown in the order condition are 1.8 times higher than the odds for showing this behavior in the disorder condition. Furthermore, the cue effect is stable when individual characteristics are included (M2). However, individuals differ in the degree to which they show other-regarding behavior. Women who walk alone are more likely than men to behave prosocially. Furthermore, older people are more likely to behave prosocially, independent of the effect of a cue. No effect was found for people with a foreign appearance.

M3 in Table 2 includes interaction terms between individual characteristics and the order cue. In M3, the main effect of a cue is not significant anymore. Further, women and men react differently to order and disorder cues in the environment, and people also behave differently when walking alone or in a group. Interestingly, in the order condition, the odds ratio for prosocial behavior of groups is higher than the odds for individuals who walk alone. One would expect that groups show systematically less prosocial behavior. Furthermore, under conditions of disorder, groups of men show the least other-regarding behavior, whereas cues do not much affect the behavior of men who walk alone. The difference between men and women is weaker in the disorder condition. In Figure 2, plots of marginal predicted means illustrate these findings.

Table 3 shows the models that include neighborhood characteristics and the interaction between neighborhood and cue. All these models control for both individual-level variables and for the different experiments.

The models in Table 3 reveal that the effect of a cue remains stable when neighborhood characteristics are added to the regression model, but effects vanish or become reversed when interaction terms are taken into account. In M1, the number of owned houses in a neighborhood is positively associated with other-regarding behavior. In M2, a cue had no effect when the interaction between the cue and neighborhood characteristics was added. The main effect of the percentage of

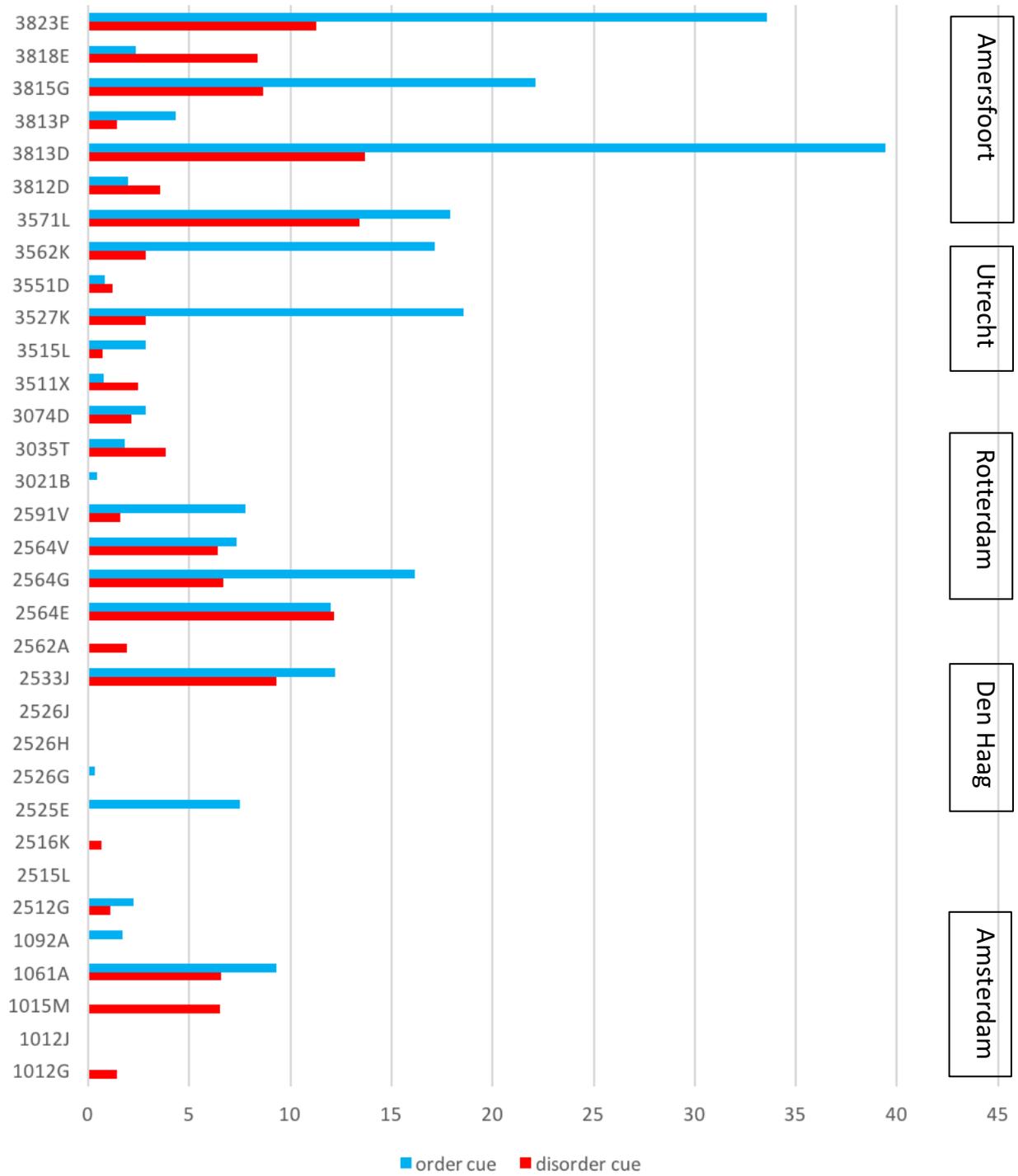


Figure 1: Other-regarding behavior (percent) by order cue, neighborhood, and city.

Table 2: Multilevel binomial regression model on other-regarding behavior for cue effect and individual characteristics (odds ratios with 95% confidence intervals in parentheses. $N = 12,528$ individuals in 33 neighborhoods).

	M1	M2	M3
Order cue	1.833 [†] (1.540–2.191)	1.830 [†] (1.528–2.192)	1.133 (0.530–2.421)
Sort experiment (reference = letter)			
Bike	0.727 (0.461–1.149)	0.727 (0.459–1.150)	0.703 (0.445–1.110)
Money	1.151 (0.740–1.786)	1.157 (0.741–1.808)	1.162 (0.742–1.817)
Individual characteristics			
Gender and group (reference = man alone)			
Woman alone		1.234* (1.011–1.506)	0.725* (0.530–0.990)
Mixed-gender group		0.730 (0.470–1.134)	0.246 [†] (0.109–0.553)
Group of men		0.659 (0.421–1.033)	0.198 [†] (0.088–0.448)
Group of women		1.269 (0.883–1.824)	0.600 (0.319–1.129)
Age (reference = below 26 years)			
26–35 years of age		1.245 (0.831–1.867)	1.343 (0.663–2.710)
36–45 years of age		2.077 [†] (1.425–3.027)	2.312* (1.203–4.443)
46–55 years of age		2.626 [†] (1.800–3.833)	2.565 [†] (1.324–4.970)
56–65 years of age		3.359 [†] (2.273–4.964)	4.156 [†] (2.134–8.091)
65+ years of age		4.000 [†] (2.607–6.136)	4.974 [†] (2.483–9.964)
Foreign appearance		0.852 (0.648–1.120)	1.020 (0.675–1.540)
Interaction with cue			
Woman alone			2.342 [†] (1.560–3.516)
Mixed group			5.663 [†] (2.177–14.731)
Group of men			7.042 [†] (2.657–18.658)
Group of women			3.307 [†] (1.526–7.167)
Age 26–35			0.908 (0.384–2.145)
Age 36–45			0.850 (0.382–1.894)
Age 46–55			1.064 (0.477–2.371)

Table 2 continued

	M1	M2	M3
Age 56–65			0.730 (0.323–1.652)
Age 65 and older			0.683 (0.286–1.631)
Foreign appearance			0.773 (0.463–1.289)
Intercept (standard error)	0.018 (0.006) [†] (0.010–0.032)	0.008 (0.003) [†] (0.004–0.016)	0.011 (0.004) [†] (0.004–0.025)
Variance (standard error)	2.622 (0.814) (1.427–4.820)	2.402 (0.754) (1.297–4.447)	2.488 (0.778) (1.348–4.593)
–2 Log likelihood	–1991.349	–1938.254	–1918.829
Intraclass coefficient (%)	44.5	42.1	43.1

* $p < 0.05$; [†] $p < 0.01$.

non-Western migrants is significant in this model, as having more non-Western migrants in a neighborhood was associated with a smaller likelihood of engaging in other-regarding behavior. However, the interaction with a cue moderates this effect in that the negative association between other-regarding behavior and the percentage of foreigners in a neighborhood becomes smaller in the order condition. Additionally, in neighborhoods with a higher number of owned properties, people are more likely to react to the order cues; thus, the effect of the order cue is present only in these neighborhoods.

M3 shows that in neighborhoods with more married people, more elderly people, and a higher neighborhood status in general, people show more other-regarding behavior. In M3, the effect of the order cue is significant, but when including the interaction terms (M4), the effect of the cue becomes negative. Moreover, the combined effect of neighborhood status with a cue is negative, indicating that the effect of the order cue depends on the status of a neighborhood in the sense that neighborhood status affects other-regarding behavior less in the order condition. In the presence of disorder cues, people in high-status neighborhoods show more other-regarding behavior. This finding can be illustrated with an anecdotic experience the observers made in such a situation. In the field experiments, one group of observers actually sometimes asked people why they put the letter in the mailbox. An older woman answered: "I saw all that litter, unbelievable what people do in these days, and I thought that nobody would see the letter. But it might be important. We are a neighborhood where we care for each other."

Furthermore, model M3 also shows that cue effects depend on the percentage of married people in the neighborhood in that the order cue is more effective in

Table 3: Multilevel binomial regression model on other-regarding behavior for cue effect and neighborhood characteristics (odds ratios with 95% confidence intervals in parentheses. $N = 12,528$ individuals in 33 neighborhoods).

	M1	M2	M3	M4
Order cue	1.829 [†] (1.527–2.190)	0.591 (0.105–3.304)	1.832 [†] (1.529–2.194)	0.328* (0.129–0.835)
Neighborhood variables				
Owned property	1.034* (1.000–1.070)	1.005 (0.667–1.158)		
Percent non-Western migrants	0.985 (0.960–1.011)	0.928 [†] (0.891–0.967)		
Residential fluctuation	1.005 (0.969–1.032)	1.005 (0.964–1.048)		
Neighborhood status			1.209 (0.997–1.466)	1.941 [†] (1.452–2.596)
Percent married			1.077 [†] (1.021–1.137)	0.967 (0.901–1.038)
Percent elderly			1.095 [†] (1.029–1.166)	1.147 [†] (1.059–1.243)
Interaction with cue				
Owned property		1.016* (1.002–1.031)		
Nonwestern migrants		1.035 [†] (1.016–1.053)		
Residential fluctuation		0.996 (0.986–1.006)		
Neighborhood status				0.754 [†] (0.667–0.853)
Percent married				1.069 [†] (1.039–1.099)
Percent elderly				0.974 (0.946–1.003)
Intercept	0.003 [†] (0.000–0.438)	0.008* (0.000–1.167)	0.003 [†] (0.000–0.001)	0.001 [†] (0.000–0.004)
Variance neighborhood level (standard error)	1.578 (0.496) (0.852–2.922)	1.604 (0.503) (0.867–2.966)	1.079 (0.351) (0.570–2.042)	1.129 (0.365) (0.598–2.131)
–2 Log likelihood	–1931.579	–1922.369	–1926.795	–1911.157
Intraclass coefficient (%)	32.4	32.7	24.6	25.8

* $p < 0.05$; $†p < 0.01$. The models are controlled for individual characteristics and type of experiment.

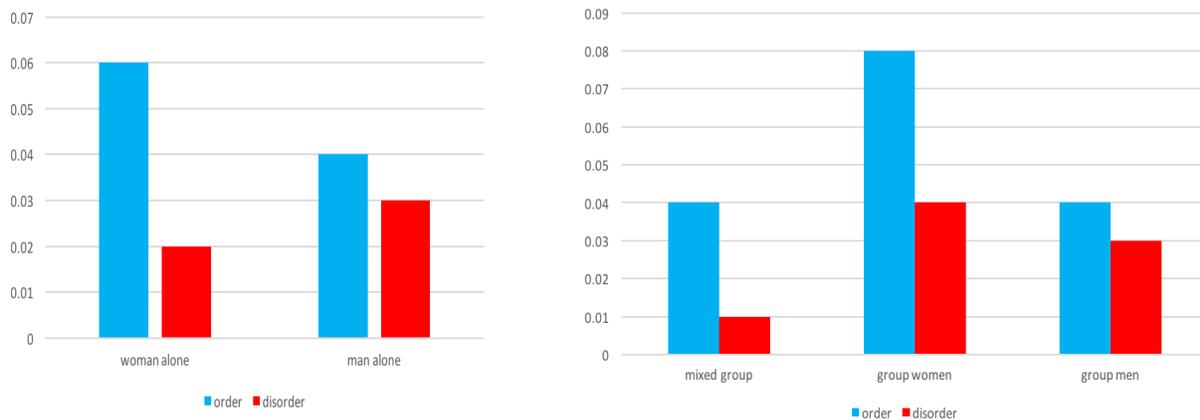


Figure 2: Marginal predicted means of other-regarding behavior for gender and group and individual in the order and disorder cue conditions.

neighborhoods with more married people. As already mentioned, more other-regarding behavior is observed in neighborhoods with more elderly people, which is in line with the finding that older individuals behave more prosocially in this regard. In a neighborhood with fewer elderly people, the difference between order and disorder cues is weaker. Finally, no differences in cue effects for neighborhoods with varying residential fluctuation are found. Marginal means plots illustrating the associations between neighborhood characteristics and other-regarding behavior are presented in Figure 3.

Table 4 presents odds ratios for key individual and neighborhood variables for the order and the disorder conditions separately. It shows that in some cases, disorder cues have a stronger association with other-regarding behavior compared with order cues and that other conditions have a greater effect in the order condition. Amongst other findings, the table illustrates once more that in high-status neighborhoods, the presence of disorder cues triggers prosocial behavior.

Conclusion and Discussion

This study considered contextual, compositional, and individual characteristics when assessing the effects of (dis)order cues. Large-scale field experiments were conducted, and information about individuals and neighborhoods has been added to the data on other-regarding behavior. By replicating and extending earlier work on cue effects for individual behavior, this study responded to a recent argument that replication deserves more attention in the social sciences (Firebaugh 2008; Watts 2017). The analysis of the field experiments led to numerous conclusions. In general, other-regarding behavior occurred in only 4.7 percent of all cases. This figure was higher in the order condition, but the absolute effect of cues was much smaller than originally thought. Although the general likelihood of other-regarding behavior was small, the analysis revealed a significant, relative, overall effect of a cue. Yet in many neighborhoods, the difference between cues was insignificant, and sometimes,

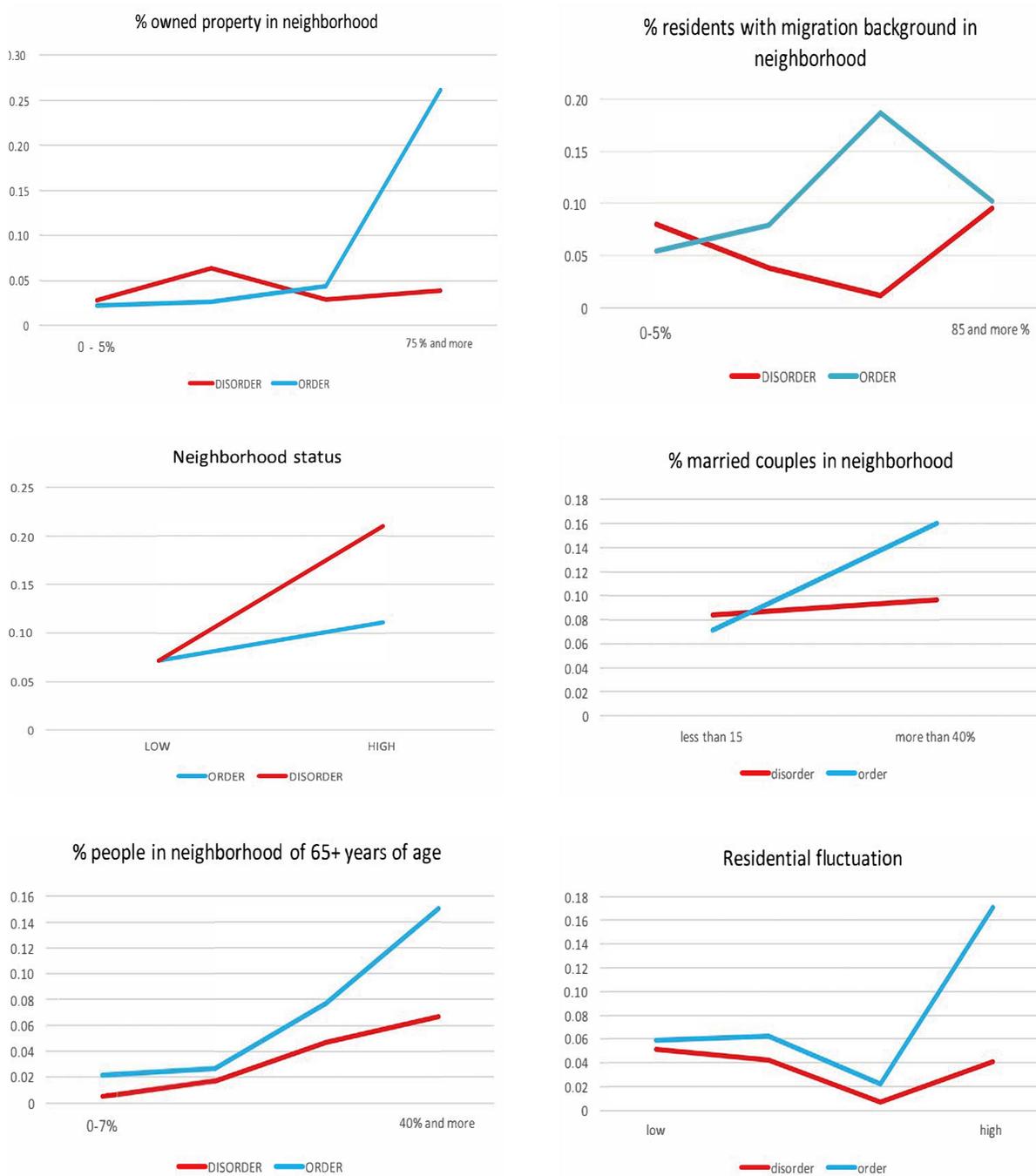


Figure 3: Marginal predicted means of other-regarding behavior for key conditions in neighborhoods and in the order and disorder cue conditions.

Table 4: Summary of odds ratios for key conditions by disorder and order cues separately.

	Odds ratios for models of disorder	Odds ratios for models of order
Individual characteristics		
Gender and group (reference: man alone)		
Woman alone	0.903	1.669 [†]
Mixed-gender group	0.333 [†]	1.516
Group of men	0.303 [†]	1.433
Group of women	0.801	2.005 [†]
Age (reference: below 26 years of age)		
26–35 years of age	1.475	1.217
36–45 years of age	2.457 [†]	2.098 [†]
46–55 years of age	2.680 [†]	2.853 [†]
56–65 years of age	3.732 [†]	3.365 [†]
65+	4.384 [†]	4.562 [†]
Neighborhood characteristics		
Percent owned property	1.012	1.055*
Percent non-Western residents	0.966 [†]	1.003
Residential fluctuation	0.991	1.003
Neighborhood status	1.334 [†]	1.066
Percent married	1.044	1.127 [†]
Percent older than 65 years of age	1.086 [†]	1.098*

* $p < 0.05$; [†] $p < 0.01$.

the effect of cues was even reversed (that is, more other-regarding behavior was shown in the disorder condition and less in the order condition). In general, the difference between neighborhoods was larger than the difference in cue conditions. Furthermore, a replication of the original Groningen letter experiment showed no cue effects at all.

These findings show that the effects of a cue or a signal depend heavily on the context in which it is presented as well as on the characteristics of the individual who reacts to the cue. The variation across neighborhoods is very important in this regard. In the multilevel models controlling for cue effect, type of experiment, neighborhood, and individual characteristics, at least 25 percent of the variation in the data was attributable to neighborhood differences. Important neighborhood composition characteristics were the number of married couples and elderly people and the status of the neighborhood in terms of education, income, and labor market position of the residents as well as the number of non-Western migrants. Additionally, contextual conditions matter, such as the number of houses owned compared with rented. When estimating neighborhood effects, it is found that in wealthier and more advantaged neighborhoods, other-regarding behavior was higher in general. In neighborhoods that are less advantaged, other-regarding behavior was lower.

In addition to the neighborhood heterogeneity, it is important to consider individual heterogeneity. Older people and women are particularly likely to exhibit more other-regarding behavior, regardless of the cue they are confronted with. Furthermore, compared with people younger than 26 years of age, all other age groups (particularly the group older than 65 years of age) showed more other-regarding behavior, independent of the cue. In general, the findings are in line with previous arguments on the contextual embeddedness of disorder (Sampson 2012).

When assessing the interaction between individual characteristics and neighborhood characteristics with the order or disorder cue, it is found that the main effect of the cue in a neighborhood often became insignificant or even reversed.

Several expectations have been formulated. First, following earlier studies, I expected that in general, cues will matter for other-regarding behavior. Second, regarding individual differences, it was expected that some people will show more other-regarding behavior because of their role, socialization, and maybe because their focus is more on the immediate environment and that the same holds for advantaged neighborhoods. Third, I argued that cue effects are moderated by contextual and compositional conditions in the neighborhood and by individual characteristics. In particular, I expected that in certain neighborhoods, the disorder cue can trigger prosocial behavior.

Regarding the first hypotheses, cue effects are weak but significant. As mentioned, they are not equal for everybody, though. Women and older people show more other-regarding behavior in general. Furthermore, in more stable and advantaged neighborhoods, people behave more prosocially. This confirms the second hypothesis about individual differences and differences between neighborhoods. Regarding the third hypothesis about the individual, contextual, and compositional moderation of cue effects, the analyses revealed that, as expected, in advantaged, richer neighborhoods (i.e., neighborhoods with a higher status), the disorder cue triggers other-regarding behavior. On the other hand, however, in neighborhoods with a high percentage of owned property, people react more prosocially in the presence of the order cue. The same is true for neighborhoods with many married couples. In neighborhoods with many non-Western foreigners, cue differences are not large, but prosocial behavior occurs slightly more in the order condition.

Moreover, when presented with a disorder cue, women and elderly people are not triggered to behave significantly more prosocially than they already do; hence, the positive association between female gender and other-regarding behavior is weaker in the disorder condition. In consequence, the evidence for the third hypotheses is mixed and calls for a deeper inquiry in further research.

The findings built upon (but are not completely in line with) the findings of Keuschnigg and Wolbring (2015), who have studied social capital effects on other-regarding behavior in the city of Munich. Like my study, Keuschnigg and Wolbring argued that in areas with low social capital, cues signaling disorder and a lack of social control provide no additional information and hence have no effect. Confirming this argument, my study found that order cues had fewer effects in alleged "low-social-capital" neighborhoods with a low percentage of married couples and people who rent property. However, Keuschnigg and Wolbring also argued that disorder effects were particularly likely to promote norm-violating behavior in

high-social-capital areas: "The deteriorating effect of social disorder is 31.7 percentage points larger in 'good' than in 'bad' neighborhoods" (p. 118). Albeit that higher levels of social capital are in general associated with lower levels of deviance (p. 122), the effects of disorder cues on norm violation are particularly strong in high-social-capital areas (p. 103). Assuming that high-status neighborhoods are also neighborhoods with high social capital and high social control, one would not expect that disorder cues in these neighborhoods would trigger other-regarding behavior (on the contrary, one would expect the opposite). However, I argued that people in these areas would be more likely to notice the disorder cue in the first place because their environments are relatively organized and that they are more inclined to establish social order because of the higher level of collective efficacy. Hence, I expected them to behave even more prosocially in the presence of a disorder cue. The finding that people in high-status neighborhoods show more other-regarding behavior in the presence of disorder confirms this idea. The differences in findings between the two studies might reflect the fact that two of the three studies of Keuschnigg and Wolbring were conducted in student dormitories and at relatively anonymous crosswalks in the city. In residential neighborhoods, people might pay more attention to collective goods in their environment and might also feel more responsible for the functioning of the neighborhoods compared with students in their dorms. A more theoretical argument in this regard is that the shadow of the future and reciprocity norms play a larger role in residential neighborhoods, particularly in neighborhoods with high collective efficacy and social capital. Notwithstanding these ideas, note that the study of Keuschnigg and Wolbring answered other research questions that did not call for the inclusion of specific contextual and individual characteristics.

My study has a number of limitations that can direct future research. First, the number of neighborhoods was relatively small. Therefore, not many predictors could be added in the multilevel analyses. Furthermore, additional knowledge about the specific neighborhoods would be important for a better understanding of contextual effects. What is going on in the neighborhoods where cue effects were reversed, and what is happening in the neighborhoods where the disorder cue had strong effects? In line with this, it is unclear why the findings of Keizer et al. (2008 and 2013) could not be established in the replication. Did the setting change within the few years between the original study and the replication? Moreover, although individuals who definitely did not see the cue were excluded from the analysis, one cannot be sure whether the others actually noticed the cue and what their thoughts were when they saw the letter. Last but not least, because other-regarding behavior is relatively rare in general, it might be hard to establish the conditions that matter.

Notwithstanding these limitations, the study contributes to the discussion about the effect of disorder cues and the broken windows theory while revealing the influence of important contextual, compositional, and individual characteristics. In this way, it also contributes to the discussion about the micro–macro link (see Alexander et al. 1987) in the social sciences.

There is no general or universal effect of the stimuli; instead, the contextual and individual characteristics moderate these effects. In neighborhoods that are more advantaged (i.e., higher status), have more married people, and the like, people

are more likely to engage in prosocial behavior. Sometimes, even disorder cues in such neighborhoods can trigger other-regarding behavior. Importantly, people in neighborhoods with a higher level of social order seem to be less perceptive of disorder cues.

The applied field experiments can be an important step in research towards designing area-based policies. The necessity for municipalities or other institutions to intervene in neighborhoods depends on their compositional and contextual characteristics. In advantaged neighborhoods, people might reinstall social order themselves, and therefore, government interventions might be less important. To conclude, the interplay between individuals and the environment influences the effect of order or disorder cues. Future research should study this interplay of the cue environment with characteristics of actors more thoroughly.

Notes

- ¹ In the famous car experiment by Zimbardo, an automobile without a license plate was parked with its hood up on a street in the Bronx, and a comparable automobile was parked on a street in Palo Alto. The car in the Bronx was attacked by passersby within few minutes. The car in Palo Alto sat untouched.
- ² Keizer et al. (2011) published a similar study testing the effect of prohibition signs. I deliberately did not refer to this study because the argument is slightly different and, more importantly, their data overlap with those of the other studies (see also Wicherts and Bakker 2014).
- ³ Keizer et al. (2014) acknowledged this criticism and provided more information about their design. However, this information did not alter the main conclusions of the Wicherts and Bakker (2014) article.
- ⁴ Negative binomial models and models using the mean rates of other-regarding behavior in a neighborhood to offset the likelihood that an individual behaves prosocially were also estimated. None of these models led to other conclusions. Furthermore, in previous analyses, the logarithm of the number of passersby in a neighborhood was used as an offset value to control for different numbers in cases in which more than one experiment was conducted in a neighborhood. Again, conclusions were not altered.

References

- Alexander, Jeffrey C., Bernhard Giesen, Richard Münch, and Neil J. Smelser. 1987. *The Micro-Macro-Link*. University of California Press. Berkeley.
- Bornstein, Robert F., and Catherine C. Craver-Lemley. 2004. "Mere Exposure Effect." Pp. 215–33 in *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgment and Memory*, edited by R. F. Pohl. Hove, UK: Psychology Press.
- Statistics Netherlands (CBS). 2016. Kencijfers wijken en buurten 2016 [Key figures and neighborhoods 2016]. Statistics Netherlands, The Hague. Retrieved September 19, 2017. <https://www.cbs.nl/nl-nl/maatwerk/2016/30/kencijfers-wijken-en-buurten-2016>.
- Chu, Doris. 2007. "Broken Windows Theory of Crime." Pp. 95–96 in *The Blackwell Encyclopedia of Sociology*, edited by George Ritzer. Blackwell Publishing. Blackwell Reference Online. <https://doi.org/10.1002/9781405165518.wbeosc147>

- Cialdini, Robert C., R. Reno Raymond, and Carl A. Kallgren. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology* 58:1015–26. <https://doi.org/10.1037/0022-3514.58.6.1015>
- De Kort, Yvonne A.W., L. Teddy McCalley, and Cees J. H. Midden. 2008. "Persuasive Trash Cans: Activation of Littering Norms by Design." *Environment and Behavior* 40:870–91. <https://doi.org/10.1177/0013916507311035>
- Firebaugh, Glenn. 2008. *Seven Rules for Social Research*. Princeton University Press, Princeton, New Jersey.
- Gladwell, Malcolm. 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. Little, Brown.
- Haley, Kevin J., and Daniel M. T. Fessler. 2005. "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game." *Evolution and Human Behavior* 26: 245–56. <https://doi.org/10.1016/j.evolhumbehav.2005.01.002>
- Harcourt, Bernhard E. 2006. "Broken Windows: New Evidence from New York City and a Five-City Social Experiment." *The University of Chicago Law Review* 73:271–320.
- Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2008. "The Spreading of Disorder." *Science* 322:1681–85. <https://doi.org/10.1126/science.1161405>
- Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2011. "The Reversal Effect of Prohibition Signs." *Group Processes and Intergroup Relations* 14:681–88. <https://doi.org/10.1177/1368430211398505>
- Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2013. "The Importance of Demonstratively Restoring Order." *PLoS ONE* 8:e65137. <https://doi.org/10.1371/journal.pone.0065137>
- Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2014. "Doing Field Studies. What Is It All About?" *Group Processes & Intergroup Relations* 17:404–10. <https://doi.org/10.1177/1368430213510750>
- Kelling, George L., and Catherine M. Coles. 1996. *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities*. New York:Free Press.
- Keuschnigg, Marc, and Tobias Wolbring. 2015. "Disorder, Social Capital and Norm Violation. Three Field Experiments on the Broken Windows Thesis." *Rationality and Society* 27:96–126.
- Knol, Frans, Jeroen Boelhouwer, and Vic Veldheer. 2012. *Statusontwikkeling van wijken in Nederland 1998–2010 [Status development of districts in the Netherlands 1998–2010]*. The Hague: Netherlands Institute for Social Research.
- Koopmans, Ruud, and Susanne Veit. 2014. "Cooperation in Ethnically Diverse Neighborhoods: A Lost-Letter Experiment." *Political Psychology* 35:379–400. <https://doi.org/10.1111/pops.12037>
- Reiter, Susan M., and William Samuel. 1980. "Littering as a Function of Prior Litter and the Presence or Absence of Prohibitive Signs." *Journal of Applied Social Psychology* 10:45–55. <https://doi.org/10.1111/j.1559-1816.1980.tb00692.x>
- Shaw, Clifford, and Henry H. McKay. 1969. *Juvenile Delinquency and Urban Areas*. Chicago: University of Chicago Press.
- Sampson, Robert J., Jeffrey D. Morenoff, and Thomas G. Gannon-Rowley. 2002. "Assessing 'Neighborhood Effects': Social Processes and New Directions in Research." *Annual Review of Sociology* 28:443–78. <https://doi.org/10.1146/annurev.soc.28.110601.141114>

- Sampson, Robert J. 2012a. "When Things Aren't What They Seem: Context and Cognition in Appearance-Based Regulation." *Harvard Law Review* 1563:97–107.
- Sampson, Robert J. 2012b. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226733883.001.0001>
- Sampson, Robert J., and Stephen W. Raudenbush 1999. "Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods." *American Journal of Sociology* 105: 603–51. <https://doi.org/10.1086/210356>
- Sampson, Robert J., and Stephen W. Raudenbush. 2004. "Seeing Disorder: Neighborhood Stigma and the Social Construction of 'Broken Windows.'" *Social Psychology Quarterly* 67: 319–42. <https://doi.org/10.1177/019027250406700401>
- Schultz, P. Wesley, Rene J. Bator, Lori Brown Large, Coral M. Bruni, and Jennifer J. Tabanico. 2011. "Littering in Context: Personal and Environmental Predictors of Littering Behavior." *Environment and Behavior* 45:35–59. <https://doi.org/10.1177/0013916511412179>
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin.
- Snijders, Tom A. B., and Roel J. Bosker. [1999] 2011. *Multilevel Analyses: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks: Sage.
- Snijders, Tom A. B. 2005. "Power and Sample Size in Multilevel Linear Models." Pp. 1570–1573 in *Encyclopedia of Statistics in Behavioral Science* (vol 3). Edited by B.S. Everitt and D.C. Howell. Manchester: Wiley. <https://doi.org/10.1002/0470013192.bsa492>
- Volker, Beate, Gerald W. Mollenhorst, Wouter Steenbeek, Veronique Schutjens, and Henk Flap. 2015. "Lost Letters in Dutch Neighborhoods. A Field Experiment on Collective Efficacy." *Social Forces* 94(3):953–74. <https://doi.org/10.1093/sf/sov106>
- Watts, Duncan. 2017. "Should Social Science Be more Solution Oriented?" *Nature Human Behavior* 1:0015.
- Weaver, Russell. 2015. "Littering in Context: A Quasi-Natural Experiment to Explore Geographic Influences on Antisocial Behavior." *Applied Geography* 57:142–153. <https://doi.org/10.1016/j.apgeog.2015.01.001>
- Wicherts, Jelte, and Marjan Bakker. 2014. "Broken Windows, Mediocre Methods and Substandard Statistics." *Group Processes & Intergroup Relations* 17:388–403. <https://doi.org/10.1177/1368430213502557>
- Wilson, James Q., and George L. Kelling. 1982. "Broken Windows: The Police and Neighborhood Safety." *The Atlantic*. Retrieved September 19, 2017. <https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/>
- Zajonc, Robert B. 1968. "Attitudinal Effects of Mere Exposure." *Journal of Personality and Social Psychology* 9:1–27. <https://doi.org/10.1037/h0025848>

Acknowledgements: This study greatly benefitted from the seminar of the program group "Institutions, Inequalities, and Life Courses" at the University of Amsterdam; the participants of the symposium "Order in Context" at Utrecht University on March 31, 2016; and from comments by Henk Flap.

Beate Volker: Department of Sociology, University of Amsterdam. E-mail: b.volker@uva.nl