Journal of Experimental Social Psychology 45 (2009) 603–613

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

"How many bad apples does it take to spoil the whole barrel?": Social exclusion and toleration for bad apples ☆

Norbert L. Kerr [a,b,*], Ann C. Rumble [c], Ernest S. Park [d], Jaap W. Ouwerkerk [e], Craig D. Parks [f], Marcello Gallucci [g], Paul A.M. van Lange [h,i]

[a] Michigan State University, 254 Psychology Building, East Lansing, MI 48824-1116, United States
[b] University of Kent, Keynes College, Canterbury, Kent CT2 7NP, United Kingdom
[c] Ohio University-Chillicothe, Bennett Hall Rm. 264, Chillicothe, Ohio 45601, United States
[d] Cleveland State University, 2121 Euclid Ave. CB 171, Cleveland, OH 44115, United States
[e] Free University of Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
[f] Washington State University, Johnson Tower 233D, Pullman, WA 99164-4820, United States
[g] University of Milano-Bicocca, viale dell'Innovazione 10 (U9), 20126 Milan, Italy
[h] Leiden University, Pieter de la Courtgebouw, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands
[i] Free University of Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 23 January 2008
Revised 27 December 2008
Available online 6 March 2009

Keywords:
Cooperation
Exclusion
Ostracism
Social dilemma

ABSTRACT

In social dilemmas, where personal welfare is in conflict with collective welfare, there are inherent incentives to act non-cooperatively. Moreover, there is evidence that the example of a few uncooperative group members ("bad apples") is more influential than the example of comparable numbers of cooperative members (a *bad apple* effect). Two studies are reported that examine the functional relationship between the number of likely bad apples and individual cooperation, and whether and when the threat of social exclusion for uncooperative behavior may effectively counter the temptation to follow the example of such "bad apples". It is shown that (a) the threat of exclusion is sufficient to counter the temptation to follow a few bad apples' example, (b) such threats cannot, however, overcome the cooperation-degrading effects of large numbers (e.g., a majority) of bad apples, and (c) the effectiveness of such threats may be greater in relatively smaller groups.

> "*Wel bet is roten appul out of hoord; Than that it rotie al the remenaunt*
> [Better take the rotten apple from the hoard, Than to let it lie to spoil the good ones there]." Cook's Tale, *The Canterbury Tales*, G. Chaucer (1380)
> "*The rotten apple spoils his companion.*" *Poor Richard's Almanac,* B. Franklin (1733)

Such proverbs reflect a rather general principle in social psychology—bad information about another person appears to have a stronger effect on our impressions, evaluations, and reactions to that person than equivalently extreme good information (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Skowronski & Carlston, 1989). In this paper we are interested in the application of this principle within social dilemmas, settings in which personal and collective welfare are in conflict (Dawes, 1980; Komorita & Parks, 1999; Messick & Brewer, 1983) and where "bad" (i.e., uncooperative) behavior is always more immediately and personally rewarding than "good" (i.e., cooperative) behavior. We are particularly interested in (1) how the uncooperative ("bad") behavior of just one or more group members may substantially reduce others' willingness to act cooperatively—or, as the proverb goes, whether "one bad apple spoils the whole barrel" – and (2) what might be done to prevent this—i.e., how and when can group members be effectively deterred from following the bad example of bad apples?

There is by now considerable and convincing evidence that group members' behavior within a social dilemma is influenced by both expectations and observations of others' behavior (e.g., Bornstein & Ben-Yossef, 1994; Braver & Barnett, 1974; Dawes, McTavish, & Shaklee, 1977; Komorita, Parks, & Hulbert, 1992; Messick et al., 1983; Schroeder, Jensen, Reed, Sullivan, & Schwab, 1983; Yamagishi & Sato, 1986). With some notable exceptions (e.g., a strong competitor's tendency to exploit uniformly cooperative others, e.g., Kelley & Stahelski, 1970), the more cooperative others in the group are (or are expected to be), the more cooperative we tend to be. However, the exact nature of the relationship between others' and our own behavior has not been well established.

☆ Portions of the reported research were presented at the 10th International Conference on Social Dilemmas, Marstrand, Sweden, August 22, 2003, and at the 7th Annual Sydney Symposium of Social Psychology, March 16–18, 2004, Sydney, Australia.

* Corresponding author. Fax: +1 517 432 2476.
E-mail address: kerr@msu.edu (N.L. Kerr).

0022-1031/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2009.02.017

Our current focus is on identifying moderating factors or boundary conditions for this relationship. One obvious moderation question is whether uncooperative group members have greater impact on our behavior than equally-extreme cooperative ones. Following Colman (1982) and Ouwerkerk, Kerr, Gallucci, and Van Lange (2005), we will refer to this as the *bad apple effect*. A number of scholars (Colman, 1982, 1995; Marwell & Schmitt, 1972; Sugden, 1984) have proposed such an effect and there is some indirect but supportive empirical evidence. For example, it has been widely observed (e.g., Andreoni, 1995; Ledyard, 1995; Pruitt & Kimmel, 1977) that with repeated play in a social dilemma, the mean rate of cooperation tends to decline. The notion that the relatively less cooperative members of the group have more impact on the group's behavioral norm than the relatively more cooperative members is quite consistent with a bad apple effect. Also, it has been reported (Messick et al., 1983; Rutte & Wilke, 1984) that providing a relatively wide distribution of false harvesting feedback in a resource-conservation dilemma leads to faster depletion of the shared resource than feedback with a narrow distribution and the same mean, just as one would expect if the extremely low cooperator had greater relative impact on others' behavior. Finally, a recent, unpublished set of studies by (Ouwerkerk, Van Lange, Gallucci, & Van Vugt, in preparation; also see Ouwerkerk et al., 2005) reports that participants were more inclined to follow the bad example of a single, relatively non-cooperative person (a bad apple) in a social dilemma than the good example of a single relatively cooperative person.

The proverb with which we began this paper, "one bad apple spoils the whole barrel", makes an even stronger claim. Not only may – as Baumeister et al. (2001) suggest – bad be stronger than good (a *bad apple effect*), but even a single bad model may also be sufficient to make the rest of the group act badly (a *one-bad-apple effect*). This possibility is supported by Kurzban, McCabe, Smith, & Wilson's (2001) findings. Using a real-time five-person game, in which participants received continuous and veridical feedback on others' current contribution decisions, they found evidence that group members strived to contribute at or slightly above the level of the person making the lowest contribution in the group (a *minimal reciprocity* rule; cf. Sugden, 1984). More direct evidence comes from a study summarized in a chapter by Rutte and Wilke (1992). They asked members of five-person groups to play a dichotomous-choice NPD game, and to begin by stating a non-binding intention. Via false feedback, they then manipulated what the other four people in the group allegedly intended to do. They had five conditions: either 0, 1, 2, 3, or all 4 of the others purportedly intended to defect. After receiving this feedback, all participants make their final and binding choices. These choices are reproduced in the solid curve of Fig. 1. As the figure shows, the function relating the number of bad apples with cooperation rate was a step-function. A single bad apple lowered the cooperation rate from about 50% to about 20%, and there was no further change in cooperation as the number of bad apples increased.

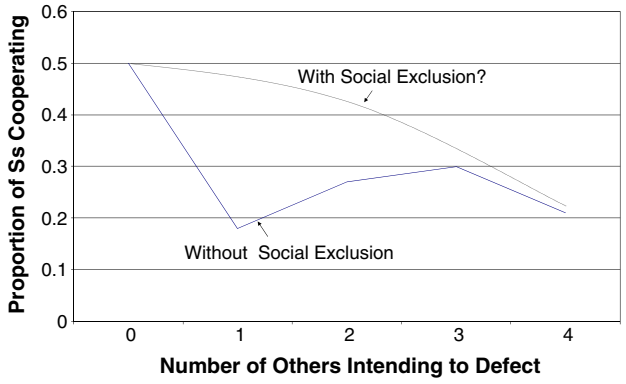

Fig. 1. Cooperation as a function of the number of "bad apples" in Rutte and Wilke (1992).

If such a one-bad-apple effect can be well established empirically (one of our present goals), it has rather disturbing implications for cooperation in human groups – namely, groups may be very vulnerable to the effects of a few or even a single uncooperative model. Social dilemmas, by definition, present clear and often strong incentives to act uncooperatively. There appears to be a non-trivial fraction of the population that will nearly always defect (Fischbacher, Gächter, & Fehr, 2001; Kurzban & Houser, 2001). It would be surprising then if most groups (especially large groups; Colman, 1995) did not include at least a few uncooperative "bad apples." Given that it is likely that there will usually be some bad apples in any group, how can groups – or at least those groups where the presence of bad apples can be detected—successfully solve social dilemmas? More generally, how did humans solve the evolutionary problem of sociality—i.e., evolve into a species that can routinely solve social dilemmas?

One generic solution to the one-bad-apple problem has been identified by evolutionary game theorists (Boyd & Richerson, 1992; Gintis, Bowles, Boyd, & Fehr, 2003; Henrich et al., 2001; Hirshleifer & Rasmussen, 1989; Kameda, Takezawa, & Hastie, 2003; Price, Cosmides, & Tooby, 2002; Sugden, 1984). It requires that two conditions be met: that humans (1) act reciprocally (i.e., cooperate and defect in response to others' actions), and (2) punish non-cooperators. Most of these models, however, are vague about the form such punishment might take. For example, in developing an evolutionary model to explain communal sharing among primitive humans, Kameda et al. (2003) refer to fighting as a way of punishing hunters who refuse to share what they kill or even those who refuse to inflict such punishments. In their laboratory work, Fehr and Gächter (2002) permitted punishment of defectors via fines. In their survey research, Price et al. (2002) assess willingness to punish defectors (e.g., draft dodgers at a time of war) via legal sanctions (e.g., prison). As you might expect in early and primarily theoretical efforts, such punishment is considered abstractly and any punishment (of equal severity or disutility) is assumed to be as effective as any other. These models do, however, suggest that such punishment will not deter widespread defection unless the punishment is costly enough for defectors and the costs of imposing the punishments not too great for the punishers (cf. Yamagishi, 1986).

If these models are correct, then a central issue in human cooperation is whether, when and how groups effectively punish uncooperative defectors. A primary objective of this paper is to examine the effectiveness of one particular type of punishment—social marginalization, ostracism (Williams, 2001), or exclusion from the group. We ask whether group members will resist the apparently strong temptation to follow the example of a few bad apples *if* by failing to do so, they risk social exclusion/marginalization in their group. It is important to note that the studies suggesting that a single bad apple can "spoil the barrel" (e.g., Kurzban et al., 2001; Rutte & Wilke, 1992), like nearly all experimental social dilemma studies, minimized the possibility of group members being able to punish one another (e.g., choices were anonymous and/or group members were led to believe that they could not interact with one another following the study).

Social psychological interest in the general effects of social exclusion (and its many variations—ostracism, rejection, bullying) has grown dramatically in the last few years (see Abrams, Hogg, & Marques, 2005; Williams, 2007; Williams, Forgas, & von Hippel, 2005, for excellent overviews). Much of this work demonstrates that social exclusion or rejection is highly aversive (Baumeister & Tice, 1990; Williams, 2001), or conversely, that there is a strong