Organizational Behavior and Human Decision Processes 120 (2013) 142–153

Contents lists available at SciVerse ScienceDirect

Organizational Behavior and Human Decision Processes

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

The price of anarchy in social dilemmas: Traditional research paradigms and new network applications

Vincent Mak[a,1], Amnon Rapoport[b,*]

[a] Cambridge Judge Business School, University of Cambridge, Trumpington Street, Cambridge CB2 1AG, United Kingdom
[b] Department of Management and Marketing, A. Gary Anderson Graduate School of Management, University of California at Riverside, Riverside, CA 92521, United States

ARTICLE INFO

Article history:
Received 7 November 2011
Accepted 16 June 2012

Keywords:
Social dilemmas
Price of anarchy
Transportation and communication networks

ABSTRACT

Research on social dilemmas has largely been concerned with *whether*, and under what conditions, selfish decisions by autonomous individuals jointly result in socially inefficient outcomes. By contrast, considerably less emphasis has been placed on the *extent* of the inefficiency in those outcomes relative to the social optimum, and how the extent of inefficiency in theory compares with what is observed in experiments or practice. In this expository article, we introduce and subsequently extend the price of anarchy (PoA), an index that originated in studies on communication in computer science, and illustrate how it can be used to characterize the extent of inefficiency in social dilemmas. A second purpose of our article is to introduce a class of social dilemmas that occur when individuals selfishly choose routes in networks, and illustrate how the concept of PoA can be helpful in studying them.

Introduction

In this paper, we introduce an index of inefficiency in social dilemmas called the price of anarchy (PoA). Our primary aim is to place the *extent* of inefficiency on the agenda of experimental research on social dilemmas, as this area of research has largely been preoccupied with *whether* socially inefficient outcomes occur and what factors influence the probability of their occurrence. To achieve our aim, we illustrate the application of PoA in a number of examples; some of these (like the classic public goods game) will be familiar to researchers on social dilemmas, while others involve route choices in networks that might be less familiar. We hope that our introductory exposition will engender new research on social dilemmas in networks.

But what are social dilemmas? In his path-breaking review of social dilemmas – a cornerstone in the development of research in this field – Dawes (1980) defined a social dilemma as an interactive decision making situation that satisfies two properties. First, the individual payoff for each agent who chooses to defect is strictly higher than the payoff for choosing to cooperate, no matter what choices are made by the other members of his group. The second property mandates that members of the population gain a higher payoff if all cooperate than if all defect. In the language of game theory, Dawes has opted to define social dilemmas as noncooperative one-shot n-person games in which (i) all the n players have strictly dominant strategies that (ii) collectively result in a socially inefficient (Nash) equilibrium.[2] Tailored to the paradigmatic n-person Prisoner's Dilemma (PD) game (e.g., Rapoport, 1970), the definition has several drawbacks. First, it does not allow for games with mixed-strategy equilibria, which are progressively more often included in the ever expanding scope of social dilemma research. Second, it unnecessarily restricts the scope of social dilemma research by excluding games with multiple equilibria.

Even more restrictive is the requirement that all the n players have strictly dominant strategies. Historically, this restriction has been relaxed in subsequent reviews of social dilemma research by Messick and Brewer (1983), Kollock (1998), and several experimental studies. In particular, Kollock offers a more general definition of social dilemmas as (p. 183): "…situations in which individual rationality leads to collective irrationality. That is, individually reasonable behavior leads to a situation in which everyone is worse off than they might have been otherwise." Kollock argues that, for example, the 2×2 Assurance game,[3] in which neither player has a dominant strategy, "is a more accurate model than the Prisoner's Dilemmas Game of many social dilemmas situations" (p. 187). And when discussing social dilemmas that satisfy the two properties listed by Dawes, he wrote (p. 185): "However, not all social dilemmas involve dominating strategies."

* Corresponding author. Fax: +1 951 827 3970.
E-mail addresses: v.mak@jbs.cam.ac.uk (V. Mak), amnon.rapoport@ucr.edu (A. Rapoport).
[1] Fax: +44 (0) 1223 339701.

[2] See the section "The price of anarchy" for a formal exposition of the idea of equilibrium.

[3] The 2×2 Assurance game is a coordination game with two equilibria: mutual cooperation, which is socially optimal, and mutual defection, which is socially inefficient.

0749-5978/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.obhdp.2012.06.006

To add another important example, and illustrate how the field of social dilemmas has been expanded and consequently enriched in the last 30 years by relaxing the definition of Dawes, consider an experiment that, ironically, was conducted by van de Kragt, Orbell, and Dawes (1983) only 3 years after Dawes had published his *Annual Review of Psychology* paper. The experiment concerns a one-shot step-level public goods game. The game is played by a group of n symmetric players with no pre-play communication. Each player is endowed with e units ($e > 0$) and must decide independently and simultaneously either to contribute all of them to the benefit of his group (i.e., the public good), or keep all of them for himself. If m or more players ($1 < m \leqslant n$) contribute their endowments,[4] then each player i receives a reward of r units ($e < r$); if $m - 1$ or fewer players contribute, then the contributors lose their endowments, whereas the non-contributors keep theirs. It is easy to verify that if exactly m players contribute their endowments, so that everyone receives the reward, while the remaining $n - m$ players (in cases when $m < n$) keep their endowments, the total group payoff is maximized. This outcome satisfies the second property listed by Dawes. However, in violation of the first property, universal contribution is no longer a dominant strategy. In fact, this "minimal contribution set" game has $n!/[m!(n-m)!]$ asymmetric equilibria in pure strategies in which exactly m players contribute. Additionally, the game has another symmetric equilibrium in which no player contributes. Therefore, while the socially optimal outcome in a step-level public goods game is an equilibrium, it is possible that a socially inefficient outcome occurs, which is also an equilibrium. Indeed, this has often been reported by experimental studies of this game or its variants (see, e.g., Chen, Au, & Komorita, 1996; Mak & Zwick, 2010; McCarter, Budescu, & Scheffran, 2011, among many others).

Our reading of the literature on social dilemmas in psychology and economics suggests that cases like the step-level public goods game are quite common. In fact, most research on social dilemmas is motivated by the observation that self-interested behavior by autonomous decision makers generally leads to inefficient outcomes; the presence of dominant strategies is not specifically required. Therefore, in the present paper, we take the broader view of defining social dilemma situations as non-cooperative games with socially inefficient outcomes. As such, social dilemmas encompass vastly different ranges of situations, from the classic context of exploitation of commons resource pools to overpopulation, deforestation, and congestion of traffic networks (as shall be described later). Common to all these social dilemmas is one single theme, namely, that there exists a "worst-case equilibrium" which is socially inefficient.

Eliciting cooperation

If individual rationality is, in general, not a sufficient condition for achieving collective rationality (e.g., Sandler, 1992), then what proposals may be advanced for eliciting behavior that increases social welfare? This question is of immense importance because of the critical role of social dilemmas in modern society. It has far reaching policy and educational implications that have been studied in much detail (see e.g., Komorita & Parks, 1994; Ostrom, 1990; Ostrom, Gardner, & Walker, 1994). Alternative solutions to social dilemmas have been reviewed and critically discussed by Dawes (1980), Messick and Brewer (1983), Van Lange, Liebrand, Messick, and Wilke (1992), Kollock (1998), and others. They include motivational solutions in which some or all of the decision makers have other-regarding (e.g., altruistic) preferences, and strategic solutions that may or may not assume changes in the fundamental structure of the game. For example, Dawes, McTavish, and Shaklee (1977) reported that pre-play discussion of the dilemma significantly reduced the frequency of socially inefficient outcomes in n-person PD games. van Dijk, de Kwaadsteniet, and De Cremer (2009) pointed out the need for common understanding among players in facilitating coordination to achieve socially optimal outcomes. Brewer (1979) and Edney (1980) suggested that cooperative solutions to social dilemmas may be facilitated by exploiting social ties arising from social group identity. Numerous studies (e.g., Isaac & Walker, 1988; Isaac, Walker, & Thomas, 1984; Rapoport & Chammah, 1965) have demonstrated experimentally that changes in the payoff structure may affect the frequency of socially efficient outcomes, such that the greater the personal return from cooperation or the lower the personal return from defection, the higher the level of cooperation. Recent research highlights the impact of uncertainty about the payoff structure, rather than just its (expected) values, on cooperation (e.g., McCarter, Rockmann, & Northcraft, 2010; van Dijk, Wit, Wilke, & Budescu, 2004). In general, these solutions shy away from calling for social designers to recommend which courses of action the players should take or, in more controversial cases, for central authorities to enforce such courses of actions. If norms of social behavior are formed over time (e.g., in small communities), then they are supposed to be established voluntarily.

Other solutions that originated in biology and computer science have taken a distinctly more authoritarian perspective. In an influential article on the "tragedy of the commons", the biologist Hardin (1968) concluded that "freedom in a commons brings ruin to all" and advocated "mutual coercion mutually agreed upon." In computer science, where it is generally not the case that agents are completely unrestricted, Roughgarden (2009) suggested that efficient joint outcomes "could be improved upon given dictatorial control over everyone's actions." Others have been looking for a middle ground between centrally enforced solutions and completely unregulated anarchy. For example, Anshelevich et al. (2008) pointed out that agents using communication networks interact with an underlying protocol that proposes a collective solution to all the users who may individually either accept or reject it; as such, the protocol designers may at least seek to promote the best possible equilibrium strategies in terms of total welfare (see the Section "Extensions and generalizations").

The price of anarchy

Imposing changes in the payoff structure, conducting pre-play communication, or establishing superordinate authority to control everyone's action, are almost always costly in terms of time and money, often infeasible, and may frequently trigger socially undesirable reactions (e.g., a negative reaction among the group members to the infringement of their individual freedom). Therefore, a key question is which proposal or combination of proposals to implement (if any), and under what conditions in order to achieve near-optimal outcomes. This question cannot be answered in practice without measuring the potential *extent* of inefficiency caused by the behavior of independent, self-interested individuals. If the extent of inefficiency, even in the worst scenario, is relatively small, then the cost of implementing procedures to elicit cooperative behavior may exceed whatever gain in efficiency that could result. But if it is relatively large, then it might be worthwhile to bring about conditions under which decentralized optimization by selfish individuals is guaranteed to produce outcomes that are near-optimal.

Three steps ought to be taken in order to answer this question. The first is to choose a formal model that defines "the outcome of selfish behavior." The second is to define a measure of the efficiency of each outcome, often referred to as a *welfare function*.

[4] If $m = 1$, this game is known as the *Volunteer's Dilemma*; it has different properties from what is discussed here.