FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 117–142
ISSN 2081–6642

Zbigniew Karpiński
Institute of Philosophy and Sociology Polish Academy of Sciences

Kinga Wysieńska-Di Carlo
Institute of Philosophy and Sociology Polish Academy of Sciences
Albert Shanker Institute, Washington, DC

Modelling Social Situations: Trust and Cooperation Among Strangers of Unequal Status[1]

Abstract

This paper presents a process of constructing and testing a new theoretical model, one which explains how status differentiation affects cooperation between partners involved in one-shot Prisoner's Dilemma (PD) situations. Bridging claims from status characteristics and collective action theories, we predict that in a PD where actors are differentiated by salient status distinctions, the rate of cooperation will vary depending on the partner's status relative to the focal actor, as well as on whether it is a simultaneous or sequential game. In order to demonstrate the challenges involved in constructing and testing new models, we describe two different one-shot, two-person PD experiments, which are designed to test our predictions. We conclude with both substantive and methodological discussion.

Key words: status, trustworthiness, Prisoner's Dilemma, theory construction

Introduction

The cumulative growth of knowledge, which is the ultimate goal of science, occurs through relating abstract theoretical claims to each other, as well as to empirical observations. The element that allows for bridging the theoretical with the empirical, for coordinating data collection, and for guiding interpretation of both, is a *theoretical model* (Skvoretz 1998). Building theoretical models that provide explanations and exact predictions of previously unexplained or loosely explained phenomena is at the core of the scientific endeavour. In this paper, we describe step-by-step the process of constructing and testing a new theoretical model explaining how salient status differences among persons with no previous history of cooperation affect their behaviour in a one-shot Prisoner's Dilemma (PD). We start our presentation with a brief discussion of some methodological issues related to theory construction and the cumulative growth of knowledge in sociology. Next, we state the substantive problem at hand and present the theories that we apply to build our

[1] Research reported in this paper was supported by a grant from National Science Centre (Grant No. UMO-2012/05/D/HS6/03118). Authors are listed in alphabetical order and contributed equally.

model and derive testable hypotheses. Finally, we report on two experiments that we designed to test our hypotheses, and discuss limitations of our designs.

Cumulative growth in sociology

Cumulative theory growth in sociology can occur through building multilevel theories (Markovsky 1997) and/or through relating theories to each other either within theoretical frameworks (Berger, Zelditch 1997) or between them (Fararo, Skvoretz 1993).

Building multilevel theories fosters our understanding of macro implications for micro processes, and vice versa. Markovsky (1997), specifying criteria for multilevel theories, notes that theories of collective action are naturally multilevel, as they focus on explaining macro level (group) outcomes as a result of the micro level (individual) choices of mutually dependent actors. Expectation states theories, which explain the emergence of status structures out of collectively- and goal-oriented encounters between individuals, and account for their impact on stratification systems in society, represent another example of a multilevel approach (Lawler et al. 1993).

Multilevel theories grow when a set of statements meets the *containment and bridging conditions*. The containment condition is satisfied when theories are comprised of statements at two or more levels of analysis (containment) and the higher level statements contain multiple lower level units, such as body/cells, structure/positions, neighbourhood/residents. The bridging condition can be met either by: constructing a conditional statement in which the level of antecedent differs from that of the consequent (e.g., if actors are differentiated by gender, then the power and prestige hierarchy of the group will reflect unequal expectations held for different genders); or by formulating definitions, wherein the subject of the higher-level statement is defined in terms of the lower-level subject, (e.g., a group consists of at least two persons). Note that the latter type of bridge is a tautology and, as such, is untestable. Constructing and testing multilevel theories entails the same procedures as developing and verifying single level theories – that is, with special attention to avoiding the reification error, and to choosing carefully methods of data analysis (Markovsky 1997).

The multidimensional model of theory growth in sociology, put forth by Berger and associates (Berger, Wagner, Zelditch 1987), focuses on yet another aspect of cumulative theory growth, namely the relationships between empirical theories within a given general framework, paradigm, or *research program*. Five different *patterns of theoretical growth* have been identified (see Berger, Wagner, Zelditch 1987; Berger, Zelditch 1997; Shelly 2002), the most relevant of which for our purposes is *integration*. Integration entails merging two or more unit theories by identifying the interrelationships between theoretical arguments with the same range and domain of application, or by specifying conditions under which the process described by alternative mechanism operate (Berger, Wagner, Zelditch 1987; Berger, Zelditch 1997; Shelly 2002).

In Berger and colleagues' model, which is an intra-program model, the concern is the growth of theories that share a common set of concepts and assumptions.

What we propose here, however, is an attempt to relate multilevel theories originating in distinct general research programs. Fararo and Skvoretz (1993) call this *integration by consolidation*.

The feedback between theory and data plays an important role in the growth of scientific knowledge. Negative empirical evaluations indicate mistakes in the theoretical structure or its empirical instantiation. Progressive revision and retesting, in most cases, are a result of the interplay between theory and its tests (Wagner 2000). Models mediate between abstract theories and empirical data (Skvoretz 1998; Wysienska, Szmatka 2002; Lovaglia, Willer 2002; Karpiński, Skvoretz 2015). Constructing a model requires a formalized (logical or mathematical) representation of the relationships between variables of interest, instantiation of abstract theoretical elements, specification of scope conditions, and identification of empirical techniques (Berger, Zelditch 1997). Below we present all of these elements.

Substantive theories: group identity and cooperation

Sociologists and other social scientists have long been interested in the effect of *diversity* on *trust* and *cooperation*. By *diversity*, we mean differentiation of members of a collectivity with respect to some salient criterion that divides that collectivity into discrete groups. *Trust* is defined in terms of a person's expectations as to whether the behaviour of another person will be benign, these expectations being based on inferences about certain personal traits and intentions of the other actor (Yamagishi, Yamagishi 1994; Molm et al., 2000). And, finally, *cooperation* refers to the collective action by members of a group to pursue common goals. Cooperative behaviour has usually been modelled using the PD, whether two-person or multiple-person (Axelrod 1984, Kollock 1998a). In its standard application, a PD is a game between self-interested players who know nothing about one another except for their incentive structures.[2] The defining features of the PD game are: (a) mutual cooperation is most advantageous to the group as a whole; but (b) individually each player is better off defecting when the partner chooses to cooperate. Questions about the effect of group differentiation on the rate of cooperation amount to asking how the incentive structure is affected by the knowledge of one's own group membership, and that of the other. From the standard game-theoretic point of view, that knowledge is supposed to have little, if any, effect on trust and cooperation, unless the players believe that group membership carries some additional information about trustworthiness or willingness to cooperate. Research finds consistently higher rates of cooperation in PD games played by members of the same group, and lower rates in games played by members of different groups, compared with a baseline setting in which players are anonymous (Brewer, Kramer, 1986; Kramer, Brewer, 1984). This result holds for real-world groups (categories), such as ethnicity or nationality (see, for instance, Yamagishi et al., 2005; Habyarimana et al., 2007), and for situations created in the laboratory, such as preference for art by Klee vs. Kandinsky (Yamagishi, Jin, Kiyonari 1999; Yamagishi, Kiyonari 2000; Aksoy 2015). Also, shared

[2] See the Appendix for a detailed description of the two-person PD game.

group membership fosters cooperation and trust in two-person as well as multiple-person games (Brewer, Kramer, 1986; Dawes, Van De Kragt, Orbell, 1988).

There are several competing theoretical arguments explaining why shared membership elevates the rate of cooperation above the level found in the baseline setting. According to social identity theory, mere difference can lead people to evaluate their in-group more favourably than the out-group (Tajfel 1982). Consequently, actors come to perceive members of their own social category as more trustworthy than members of another social category (or categories). Acting on this perception, they are more likely to cooperate with an in-group member than with an out-group member (Brewer 1979; Brewer, Kramer 1986; Kramer, Brewer 1984). Research by Toshio Yamagishi and his collaborators, however, challenged this explanation by showing that: (a) shared membership is conducive to cooperation in simultaneous but not in sequential games, in which the expectation of direct reciprocity is a stronger predictor of cooperation (Yamagishi, Kiyonari 2000); and (b) shared membership promotes cooperation only when such sharing is common knowledge (Yamagishi, Jin, Kiyonari 1999). In other words, in games played by members of the same group, both players have to be aware of each other's group membership, since, otherwise, their common identity will not be sufficient to overcome the temptation to defect unilaterally. To illustrate, suppose two players, A and B, play against one another. Both belong to the same group, but only A is aware of that fact. Player A also knows that B is not aware of their common group membership. Because B is not aware that A is his or her fellow group member, B has no special incentive to treat A favourably. Expecting that, A also will not behave favourably towards B. As a result, the rate of trust and cooperation will be about the same as in the baseline setting. This is where Yamagishi's work departs from social identity theory, since the latter would predict that A – being aware of his or her common group identity – would choose to cooperate with B. In other words, social identity theory claims that actors cooperate with in-group members because they are in-group members, whereas Yamagishi claims that actors cooperate with their in-group members because they expect their in-group members to cooperate with them, but also that the expectation of cooperation has to be mutual, which only happens if both players are aware of their common identity. Research provides stronger support for Yamagishi's proposition than for social identity theory (Foddy, Yamagishi 2009; Yamagishi et al., 2005; Yamagishi, Kiyonari, 2000; Yamagishi, Jin, Kiyonari, 1999; Yamagishi, Mifune 2008).

Recently, Brent Simpson (2006) attempted to reformulate the social-identity theoretic account of cooperation by combining concepts from social identity theory (Tajfel 1982, Turner 1985) with the fear-and-greed approach to the study of social dilemmas (Ahn et al. 2001; Dawes et al. 1986; Rapaport, Chammah 1965; Simpson 2003). According to this approach, defection in a PD can be motivated by fear or greed, the former meaning the motivation to avoid being exploited by a non-cooperative partner, and the latter being the temptation to free ride on the partner's cooperation. Both of these motivations are defined in terms of game parameters (see the Appendix for a more formal treatment). More specifically, greed is equal to the difference between the payoff to unilateral defection (T) and the reward for cooperation (R). Fear, in turn, is measured as the difference between the payoff

for mutual defection (P) and unilateral cooperation (S). In other words, the larger is the payoff for unilateral defection relative to the reward for mutual cooperation, the greater one's greed, or the inclination to exploit the other. Similarly, the larger is the payoff for mutual defection in comparison with that for unilateral cooperation, the more afraid one is of the other's exploitation. By manipulating the payoffs, one can change the amount of fear and greed in the game, leading to more or less cooperation, depending on which payoffs are changed and in which direction.

Simpson couples the concepts of fear and greed with a central notion of social identity theory, the metacontrast principle. According to this principle, behavioural responses to the distinction between the in-group and the out-group are driven by: (a) maximization of inter-group differences; and (b) minimization of intra-group differences (Turner 1985, Hogg 1996). When a player is motivated by greed, cooperation with a fellow group member achieves these goals. But, when a player is motivated by fear, cooperation realizes the former goal, whereas defection realizes the latter. Thus, the two goals cancel each other out, and, consequently, we have no reason to expect any effect of identity on responses to the fear component of social dilemmas. This reasoning leads Simpson to propose that when players in a prisoner's dilemma make their decisions sequentially: (a) the player who moves first will not be affected by common identity, as the first decider in a sequential PD is motivated only by fear; but (b) the player who makes the second move will be affected by common identity, as he or she is motivated only by greed. This is where the fear-and-greed approach differs from Yamagishi's because Yamagishi proposed that common identity has no effect on cooperation in sequential games (Yamagishi, Kiyonari 2000). Research provides strong support for Simpson's reasoning (Simpson 2006).

To sum up, the three perspectives – social identity theory, Yamagishi's generalized reciprocity hypothesis, and Simpson's fear-and-greed approach: (a) agree that shared group membership elevates the rate of cooperation above the levels observed in anonymous settings, when decisions are made simultaneously; but (b) differ with respect to predictions concerning sequential games, with social identity theory claiming that common identity will increase cooperation in both simultaneous and sequential games, Yamagishi claiming it will have no effect in sequential game, and Simpson predicting that it will only affect the decision made by the second decider.

Substantive theories: Status and cooperation in PD

All the theories discussed in the preceding section are limited to the study of cooperation when the distinction between the in-group and the out-group is based on a *nominal* characteristic. However, characteristics that differentiate social actors may have not only nominal, but also *status* value (Berger, Webster 2006; Ridgeway 2001, 2014). Status distinctions are different from nominal distinctions in one important respect: while the latter give rise to in-group bias, i.e., preferential evaluation and treatment given to members of one's own group, regardless of the group's position in a social hierarchy, status distinctions give rise to *status beliefs* that accord more competence, esteem, and general social worth to one 'state' of the underlying

characteristic than to other state(s) (Ridgeway 2006). In addition, both high- and low-status actors accept the status beliefs.

Status distinctions can be argued to have an important implication for cooperation in social dilemmas. High-status actors are perceived as more competent, respected, and esteemed members of the collectivity, and also as more *group-oriented* and *cooperative* (Ridgeway 1982). Consequently, high-status actors can be expected to cooperate more in social dilemmas, compared with low-status actors. Thus, playing a PD game against a high-status actor reduces one's fear (of being exploited), but not necessarily one's greed (to take advantage of the other).

Choices made by high-status actors are not only more likely to be more cooperative, they are also more likely to be perceived as *normative*, thus providing low-status actors with cues to 'desirable' or 'appropriate' behaviour in a given situation. As a result, low-status actors will follow the behaviour of the high status actor due to the former's belief that it is the right thing to do. High-status actors are also more influential, so they can choose to cooperate expecting to influence low-status members (Sell 1997). Moreover, high-status actors experience positive emotions, (Lovaglia, Houser 1996; Lovaglia 1997) and positive emotions foster integration (Kemper 1987, 1991). Positive emotions induce behaviour that ties group members together, whereas negative emotions tend to draw group members away from each other. Positive emotions experienced by high-status members may therefore compel high-status actors to act in a manner that binds low-status actors to the group (Lovaglia, Houser 1996).[3]

To sum up, the reasoning presented above asserts that when actors who differ in status play a sequential PD game, the level of cooperation will depend on who initiates the game. If the high-status player makes the first move, his or her choice is likely to give the low-status player a cue as to what he or she should do himself or herself. So, if the high-status actor initiates the game by acting cooperatively – which we expect him or her to do, given the relationship between status and group orientation – the low-status player is more likely to respond in kind than he or she would respond if the information about status was unavailable. Furthermore, if the greater likelihood of cooperative behaviour on the part of the low-status actor reflects his or her belief that cooperation is the correct behaviour because this is what the high-status actor does, then one can expect the low-status actor to respond cooperatively to the behaviour of the high-status player, and to act cooperatively in future interactions involving different people.

On the other hand, the player who makes the first move in a sequential PD game is motivated by fear, while the second player is motivated by greed. As a result, the high-status actors' willingness to cooperate in a sequential PD played against a low-status actor can be reduced by the former's fear of being exploited. The high-status actor's fear can be further exacerbated by the expectation of the low-status actor's

[3] It is beyond the scope of this paper to review existing research on the link between status and contributions to public goods, but it should be noted that it demonstrates both that contributions to public good lead to higher status (e.g., Willer 2009), as well as that higher status actors, given a chance, initiate contribution more often than low status actors and they contribute more. Low status actors, in return, contribute more when following the lead of high status ones (e.g., Sell 1997; Kumru, Vesterlund 2010; Simpson, Willer, Ridgeway 2012).

being a less cooperative group member. In other words, the low-status actor is motivated by greed, and this motivation is even stronger than it would be in a situation where players had no information about their relative statuses. Consequently, the high-status actor has little incentive to make a cooperative move.

When a low-status actor initiates the game, however, the situation changes considerably. As a first decider, the low-status actor faces the fear of being exploited, but this fear is reduced by the knowledge that the partner is of higher status, and therefore is likely to respond cooperatively. Thus, the low-status actor has reason to expect the high-status actor to respond in kind to his or her cooperative move, which induces him or her to choose cooperation. This reasoning leads us to the following proposition:

> *Multilevel Hypothesis* 1: In a sequential PD game, if actors are differentiated by a salient status characteristic, the rate of cooperation will be higher when the low-status actor initiates the game than it is when the high-status actor does so.

Actors are predicted to adjust their behaviour in a PD game based on the information about their partner's status. Irrespective of whether the actor is of high or low status, knowing that his or her partner is of high status encourages the actor to behave cooperatively, expecting the partner to respond in kind. When the actor knows the partner to be of low status, he or she is less likely to cooperate, expecting the partner to behave selfishly. That is, actors are more likely to cooperate in the PD game when their partner is of high status than when he or she is of the low status. Thus, our next hypothesis can be stated as follows:

> *Multilevel Hypothesis* 2: In a sequential PD game, the rate of cooperation will be higher when the partner is of high status than when he or she is of low-status.

This prediction can also be presented in a somewhat different way. Let HH denote a pair of players, each of whom has a high status, and let HL represent a pair of players of which the first player has a high status, and the second player low status. Then, according to hypothesis 2, the rate of cooperation in pairs of type HH is predicted to be higher than in pairs HL. Similarly, we expect the rate of cooperation in pairs of type LH to be higher than in pairs LL.

We can also expect a higher rate of cooperation when both players are of high-status and a lower rate of cooperation when they are both of low status. Note that when both players are high-status, each can expect the other to behave in a cooperative manner, and so each has an incentive to act cooperatively as well. Similarly, when both players have low statuses, each has little motivation to choose cooperation, as each believes their partner will not cooperate. Therefore:

> *Multilevel Hypothesis* 3: In a sequential PD game, the cooperation rate among high-status actors will be higher than that among low-status actors.

Arguably, hypotheses 1 through 3 extend to simultaneous games as well. First, the type of game does not affect the nature of the expectations concerning the other's cooperativeness. All else being equal, the overall rate of cooperation may be

somewhat different in sequential games than in simultaneous ones, but the effect of status can still be said to operate in the same way in both types of games. As stated above, both fear- and greed-based motivations are present when actors make decisions in simultaneous games. Still, when status is salient in a prisoner's dilemma, high-status actors are more sensitive to the fear component, whereas low-status actors are more attuned to the greed component. This means their motivations are very similar to those in a sequential game in which the first player is driven solely by fear and the second solely by greed. Thus, introducing status into the situation can be expected to reduce differences in the rate of cooperation across simultaneous and sequential games. Note also that in simultaneous games both actors respond to expectations about the other's behaviour, whereas in sequential games this is true only of the first deciders, as the second deciders respond to actual behaviour (Yamagishi, Kiyonari, 2000). Insofar as, in the experiment that we report below, we focus on the behaviour of the first deciders, this is yet another reason to expect little difference in the effect of status on cooperation across sequential and simultaneous games.

Let us present a summary of our predictions. With two levels of the subject's status and two levels of the partner's status, there are four possible combinations. Let us use L to denote low status and let us use H to denote high status. The four combinations are HH, HL, LH, LL, with the first symbol in each pair referring to the subject's status. Thus, for instance, HH refers to pairs in which both the subject and the partner are of high status, HL means pairs in which the subject has high status and the partner has low status, and so forth. Finally, we use the symbol '>' to indicate the ordering of pairs in terms of the expected cooperation rate, such that HH>LL indicates that pairs in which both players are of high status are expected to have higher cooperation rates than pairs in which both are of low status.

Table 1. Summary of theoretical predictions

Hypotheses	Game type	Prediction
Hypothesis 1	Simultaneous and Sequential	LH>HL
Hypothesis 2	Simultaneous and Sequential	HH>HL, LH>LL
Hypothesis 3	Simultaneous and Sequential	HL>HH

Scope conditions

Particular theories apply to certain kinds of phenomena under certain conditions. *Scope conditions* specify the circumstances under which the relationships expressed in theoretical hypotheses are expected to hold true (Foschi 1997; Cohen 1980, 1989; Markovsky 1994; Walker, Cohen 1985). In other words, they are abstract and universal statements that define properties that must be present or absent in a situation in order for the theory to be applicable, testable, or assumed to be true. Scope statements (also referred to as boundary conditions, see Shelly 2002) assert nothing about the truth of theoretical principles. The importance of the scope

statements is simply that wherever scope conditions are met, a theoretical claim is applicable. Nothing in either the scope statement or the theoretical argument guarantees that any particular situation will meet the scope conditions (Cohen 1989, p. 83). Efforts to integrate two or more theories will involve a process of rectifying two or more sets of scope conditions (Shelly 2002).

Expectation states theories are limited in their applicability to goal-oriented groups in which group members value the task and share the motivation to achieve success and avoid failure, where standards for evaluating success and failure are at least partly under the control of group members, and where group members recognize that they are interdependent and thus must take into account other group members' actions. In addition, status characteristics theory specifies that group members must believe that some characteristics are instrumental to achieving the group's goal – that is, possessing a high state of the characteristic increases the likelihood of success. Collective action theories, on the other hand, apply to mixed-motive situations, which are situations in which there is an incentive to defect despite the fact that mutual cooperation is to the group's benefit (Sell 1997).

It appears that the two theoretical frameworks differ regarding the situations to which they apply, although these situations are not mutually exclusive. As actors are interdependent, they need to take each other's (potential) actions and motivations into account. As there is a mutually satisfying outcome that is better than when no cooperation occurs, and that is easily identifiable, we can also assume that task orientation is present, and that there are characteristics instrumental to achieving it. As our primary goal is to explain status effects in PD situations, the hypotheses summarized in Table 1 apply to situations in which actors are differentiated with respect to status and face a social dilemma. In the simplest case, the situation is comprised of only two actors who may be of the same or different status. Status is, of course, not the only dimension of differentiation in social groups. Distinctions other than status have powerful effects on cooperation in social dilemmas, and these distinctions can interact with status. The focus on status in our work is not to imply that we believe categorical distinctions and social identity effects to be unimportant. Rather, we propose to *abstract* status from other dimensions of differentiation to study its effect in *isolation* from them.

The experiment

In order to test our predictions, we needed subjects to play either sequential or simultaneous PD games in pairs that are either of unequal or equal status. Also, the latter pairs are further divided into those made up of high-status players and those comprised of low-status players.

There are three variables that we explicitly manipulated: subject's status (high vs. low); partner's status (high vs. low); and the type of the game (sequential vs. simultaneous). This yields a total of 8 treatments. The design of our experiment combines the standard experimental setting used in expectation states research (see Berger 2007 for a review of the setting) with a two-person investment game

setting (see Berg, Dickhaut, McCabe 1995; see also Sell 2007 for a review of social dilemma experiments in the social sciences).

The expectation-states experimental setting was employed to induce status differences between our subjects. To study the effect of these differences on the rate of cooperation, we use the investment game, which has the structure of a PD game, and has been used in many studies as an experimental 'operationalisation' of a PD game (see Simpson 2006; Yamagishi, Jin, Kiyonari 1999 for examples). Given our hypotheses and research interests, our focus in this experiment was on the investment game. In other words, because our hypotheses are not concerned with the emergence and reproduction of status hierarchies in task groups – a major theme in expectation states theory – we used the expectation-states setting instrumentally, in order to introduce status distinctions into our experimental situation. Beyond that, we did not make any substantive use of data generated in the part of our experiment that employed this setting.

The first part of the experiment: assigning ability

In order to test our hypotheses, we needed to introduce *status differences* into our experimental situation. In status characteristics theory, status differences determine levels of assigned competence and corresponding expectations regarding an actor's performance at a group task. In many social-scientific experiments investigating links between status and collective action, status differences are introduced using quasi-experimental measures, such as level of education (e.g., Simpson, Willer, Ridgeway 2012), gender (e.g., Sell 1997), or scores on a quiz administered at the beginning of the study (List 2007; List, Cherry 2008; Kumru, Vesterlund 2010). Using these characteristics as measures of status in an experiment is relatively easy, does not require cover stories, and usually avoids deception. The drawback, however, is that some of these characteristics, such as gender, give rise to social identity effects, and identity has been found to have a powerful effect on cooperation. Also, the quasi-experimental measures can be nontrivially associated with the dependent variable, i.e., cooperation, reducing the internal validity of the experiment.

To avoid such problems and study the 'pure' status effect, we decided to base our measure of status on the *standard experimental situation* in expectation states research (see Troyer 2002, Berger 2007 for an overview). In expectation states experiments, subjects first work individually on a series of problems that are said to involve a recently discovered perceptual or cognitive ability (named 'contrast sensitivity ability' and 'meaning insight ability,' respectively). Subjects are told that, according to collected evidence, these abilities have no association with other skills, abilities, or personal attributes, but other properties of these new skills have yet to be investigated. The problems are quite ambiguous, but the subjects are explicitly told that there is a correct answer and individuals that are more contrast-sensitive or have more meaning insight ability are more likely to answer correctly. In our experiment, we used a version of a setting designed by Lisa Troyer (2002), which utilizes visual problems related to 'contrast sensitivity.'

Once a subject completes all his or her tasks, he or she is assigned a 'score,' which is supposed to be equal to the number of correct answers he or she gave. The scores are either 'high' or 'low,' indicating their ability (or lack thereof) to solve most of the problems correctly. Contrast sensitivity is actually a fictitious ability and the tasks given to the subjects do not have a correct solution in any meaningful sense. The 'scores' are randomly assigned to the study participants, regardless of their responses to the problems. Thus, the classification of the subjects as having a high or low level of the ability is independent of their actual abilities or characteristics. The sole purpose of the first part of the experiment is to lead the subjects to believe that: (a) there is such an ability as contrast sensitivity; (b) the problems they just solved measure it in a valid way; and (c) they possess either a high or low level of the ability. With these beliefs, they proceed to the second part of the study, in which they are paired with a partner to work as a team on another set of contrast sensitivity problems.

The second part of the study: the link between status characteristic and cooperation

Two features of the second part of the experiment are crucial. First, the subjects know their own contrast-sensitivity score, as well as that of their ostensible partner. Second, as they are supposed to work as a team and take one another's propositions into account, they are given the opportunity to exchange their suggested answers to the task at hand before making their final decisions. At each trial, a subject first submits his or her initial solution and learns the initial solution proposed by his or her partner. If their propositions are different, the subject can either stay with his or her initial idea, or discard it and accept the alternative proposed by the partner. In expectation states theory, the probability of staying with the initial response is interpreted in terms of *rejection of influence* and modelled as a function of differences in performance expectations linked to status differences between team members. As mentioned earlier, we were not really interested in studying patterns of influence. Instead, we used the second part of the experiment to induce a link between status differences and cooperation by providing subjects with explicit information that their partner did or did not act as a 'good team player,' depending on the partner's status. More specifically, in line with the research by Ridgeway (1982) in which high-status actors were found to be more group-oriented and cooperative, the high-status partners were described as 'good team players' and the low-status partners as failing to act as such.

The 'partner' in the second part of the study was actually simulated. Each subject worked independently of the others, although they were led to believe that they collaborated in pairs. The 'partner' was *pre-programmed to disagree* with the subject on an initial solution for a specified percentage of the time.[4] The trials in which the 'partner' agreed with the subject were randomly distributed. The reason for using a simulated rather than a real partner is that it permits the researcher to manipulate explicitly the frequency of disagreements between the subject and the

[4] In expectation states experiments, the percentage is 75. In our study, we set the percentage to 80 per cent in our first experiment and to 70 in the second.

partner, and it is precisely the study of what happens when they disagree that has been at the centre of much research within the expectation-states tradition. Given that the objective of our study was not to test predictions derived from expectation states theory, we could let the subjects interact over the computer network with real rather than simulated partners. We decided to follow the original procedure, however, because it gave us the opportunity to observe if we were able to replicate the results of earlier studies on status characteristics. This is an important consideration. If we failed to *replicate* those results, it would imply that there is something wrong with our experimental procedures, instructions, or protocols.

The third part of the study: investment game under status differentiation

After the second part of the study came to an end, the subjects were asked to complete a short survey. They then received the feedback on their partner's group orientation. At this point, they were informed that the study had come to an end and they could go, but also that we were giving them an opportunity to take part in another, apparently unrelated experiment, which would allow them to earn extra money. Before they made their decision as to their participation in that additional experiment, however, we informed them about its objective and procedures.

The instructions that followed presented the new experiment as one in which the subjects are again required to work in pairs. It was emphasized, however, that they might have a different partner in the new experiment than they had previously. We did not want our subjects to think they would be paired with the same person, because the history of their relationship might have distorted the subjects' behaviour. Each member of a pair then received a pool of 'tokens' from the experimenter and had to decide whether to keep the endowment to him or herself, or divide it. Subjects were informed that the experimenter would double each token given to their partner. Participants could therefore benefit greatly by exchanging tokens – when one gave all his or her tokens to their partner and vice versa, both would end up having twice as much as at the beginning of the game. Knowing this, however, each of them was also tempted not to act in this manner because when one gave all his or her tokens and the other gave none, the latter would end up having three times what they started with and the former with nothing. As a team, participants were best off cooperating with one another, with cooperation meaning here the transferring of one's endowment to one's partner, but individually each benefited by unilaterally defecting, i.e., failing to share their resources with their partner. Unlike in the 'standard' PD game, the decision was not binary (cooperate or defect), as the subjects could give all or none of their tokens to the partner, or any amount in between.

Our instructions made it clear to the subjects that the decision they would face in the game was not a trivial one, as it would affect their payment, and that the tokens had a monetary value. Giving tokens to a trustworthy partner may be highly beneficial, because trustworthy partners are likely to honour trust placed in them and reciprocate, but misplaced trust may cost the subject dearly. The instructions

then stated that, to help them make their decisions, the information about their 'new' partner's performance in the earlier experiment would be displayed on screen, before they chose to participate in the game. Importantly, the performance of the 'new' partner matched closely that of their previous partner, as we wanted the subjects to 'project' onto their 'new' partner the link between their partner's scores and his or her willingness to cooperate.

One element of the instructions was varied randomly – namely, the information about the type of game. That is, half of our sample was informed they would be making their decision *simultaneously* with their partner, so that they could learn what the partner did only after they had made their own decision. The remaining subjects were informed that they would make their decisions *in sequence*, meaning that the first mover has no way of knowing their partner's decision, but the second mover makes his or her decision knowing what the partner did. It is, however, important to keep in mind that in sequential games the subject was always assigned the role of the first decider, although they were led to believe that this was due to a random choice. After the game-type manipulation, the subjects proceeded to two practice trials.[5] Once the practice trials were over, the subjects were matched with a potential partner, and could see the information about the partner's performance in the previous experiment. They then made a final decision about participation in the game. When they chose to participate, the game commenced. Note, again, that the partner was actually simulated and programmed to match the choice made by the subject, so that the subjects who gave nothing to their 'partner' received nothing and those who gave all their tokens received their 'partner's' whole endowment. When the game was completed, the subjects were asked to fill in a short survey and then summary information with their earnings was displayed on the screen.

Study 1

The first study was conducted in May and June 2014. The participants were students of public and private universities in Warsaw, Poland. They were recruited using email announcements, ads placed on social networking sites, and posters distributed in their schools. These ads and announcements emphasized the opportunity to earn money by taking part in a study of 'group decision making.'

After arriving in the laboratory, subjects were seated at computer stations. Next to each station was an informed consent form, which the subjects were asked to read and sign once they had gone through all the experimental instructions that were to be displayed on the screen. The experimenter also reminded them that they could leave the laboratory whenever they chose to do so, regardless of whether or not they completed the whole experiment, and without having to provide any reason. They were also told that leaving the experiment early would have no consequences for them except for its reducing the amount they would be paid, as their payment was dependent upon their performance in the experiment, and that leaving early could result in their

[5] In the sequential-game treatments, the subjects were the first deciders in one practice trial and the second deciders in the other.

receiving no payment at all. After that, they were asked to begin, and the experiment proceeded as described in the preceding section. However, a few details bear mentioning.

First, at the beginning of the session, the subjects were asked to provide information about their gender, age, and their year of study. Second, before the start of the second part of the experiment, in which subjects solved contrast-sensitivity problems in pairs, information about gender, age, and the year of study of their partner was displayed, along with the partner's score from the first part of the study. In order to control for the effects of these characteristics, the 'partner' was always of the same gender and a similar age as that of the subject. Furthermore, the subjects were given 20 contrast sensitivity tasks in the first part of the study and another 20 such problems in the second part, when they were matched with a partner. Thus, each subject was given as many as 40 problems in the course of the experiment. Finally, the information about age and gender was not displayed when an ostensibly new partner was assigned in the third part of the experiment.

A total of 118 subjects took part in the study. Each experimental session was comprised of an even number of participants, between four and eight, all of whom were in the same room, so that each of them could see how many other participants were present. Having an even number of subjects was necessary to validate the manipulation that they were working in pairs during the second and the third part of the study. When one of the participants failed to show up, a research assistant sat in his or her place. Also, because, as indicated above, at the beginning of the second part of the study, the subjects were informed that they were paired with a person of the same gender, there had to be at least two male participants and two female participants in each session to validate that information. After the second part, the subjects were informed that their remuneration was equal to 30 PLN. They could increase it by taking part in an additional study, in which their task would be to allocate 100 'tokens' between themselves and a partner, with each token worth 0.05 PLN. Thus, the tokens received from the experimenter were worth 5 PLN. Mutual cooperation could double the amount.

At the end of each session, the experimenter revealed all the deceptions used in the experiment, gave a detailed reason for why the deceptions were used, and responded to questions and comments from the subjects. Each session lasted about 40 minutes.

To sum up:

In the first part of the study, subjects solved 20 contrast-sensitivity tasks, after which they were randomly assigned either a high or low score;

In the second part, they were paired with a partner of the same gender and a similar age to work in pairs on another set of 20 contrast sensitivity tasks. The information on the partner's status was randomly varied. For the first two parts of the study, they earned 30 PLN (roughly, 10 USD at the time the study was conducted);

In the third part, they played an investment game by allocating 100 tokens between themselves and their partner, ostensibly a different person than their partner in the second part. Their new partner had similar contrast-sensitivity scores as the previous partner, but was of unspecified gender, year of study, and age. The tokens were worth 0.05 PLN each, so the endowment received from the experimenter amounted to 5 PLN.

Results of Study 1

Analyses of data from the second part of our study show that we were able to reproduce the results of earlier experiments using the standardized setting of the expectation states theory. That is, we observed that $P(S)$, or the probability of subjects staying with their initial decision was the highest for pairs HL – that is, situations in which a high-status subject was paired with a low-status partner – and it was the lowest for pairs LH. For status-equal treatment the probabilities were close to one-half. One-way ANOVA for $P(S)$ gives $F_{3,114} = 22.08$. The result is highly significant, well beyond the conventional significance levels. Thus, our status manipulation can be said to have been effective.

In Table 2, we present means and standard deviations of the cooperation rates across the conditions of our experiment. The results turn out to be inconsistent with what we predicted. First, the means and medians for pairs HL and LH are different, but the direction of the difference contradicts Hypothesis 1, as the average for the former pair is higher than that for the latter; also, the difference is not statistically significant ($p = 0.21$).[6]

Table 2. Descriptive statistics for the cooperation rate in Experiment 1

Game type	Subject's status	Partner's status	N	Median	Mean	SD
Sequential	H	H	17	.860	.838	.187
Sequential	H	L	13	.850	.754	.295
Sequential	L	H	13	.550	.636	.198
Sequential	L	L	13	.600	.669	.211
Simultaneous	H	H	13	.800	.719	.284
Simultaneous	H	L	18	.850	.761	.274
Simultaneous	L	H	19	1.000	.817	.231
Simultaneous	L	L	10	.900	.845	.171

For simultaneous games, the result is similar: although the difference in medians is larger, the means for pairs HL and LH are quite similar and not significantly different (p=0.55). Second, the means for pairs HH are somewhat greater than those for pairs HL, in line with hypothesis 2, but that difference is most likely due to an outlier, since the medians are almost identical. Also, the difference in means is not significant (p=0.196). As for the low-status subjects, the means and medians for pairs LH and LL are very similar (p=0.71). Consequently, we have no evidence to support Hypothesis 2, which predicts that subjects modify their behaviour in the investment game depending on the partner's status. As we can see, high-status

[6] The reported p values come from a series of significance tests based on linear combinations of coefficients from a linear regression model, with cooperation as the dependent variable and status and game type manipulations as the independent variables.

subjects cooperate at a relatively high-rate, regardless of their partner's status, whereas low-status subjects cooperate at a relatively low rate, regardless of their partner's status. Thus, as regards sequential games, some status effect is observed, but in a direction opposite of that we predicted. Interestingly, the results are quite different for the simultaneous games. As we can see, it is the low-status subjects that appear to cooperate more in simultaneous games. That is, the means for pairs LH and LL are higher than the corresponding means for pairs HH and HL. However, the differences are small and not significant (the HH-HL difference: p=0.81; the LH-LL difference: p=0.47).

As for the third hypothesis, the difference in the cooperation rate in sequential games between pairs HH and LL is quite large, positive and marginally significant (p=0.056). This is consistent with our prediction, but we failed to find a similar effect in simultaneous games (p=0.18).

In general, there are two possible explanations of our not finding support for our predictions: either the predictions were wrong or we failed to reproduce essential features of our theoretical model. The latter interpretation did not require of us to abandon or discard the hypotheses and the reasoning on which they were based, so we decided to consider this interpretation first. Considering flaws in our experimental design was also justified insofar as alternative hypotheses implying higher cooperation rates in both simultaneous games (because of positive emotions) and sequential games (because of high status actors influence) were only partially consistent with our observations. Recall that, even though high status subjects cooperated more than low status ones, regardless of who the partner was in the sequential games, there were no differences across status conditions in the simultaneous games.

We have identified several features of our study design that may have contributed to an unusually high cooperation rate and our inability to find the hypothesized status effects. First, we found a substantial and significant gender effect: male subjects turned out to have cooperated at a greater rate than female subjects. In comparison with otherwise identical women, men gave their partners about 14 tokens more, on average (p = 0.003). This was quite perplexing, as it suggested that gender was a salient distinction in our design, contrary to our intentions. Recall from our description of the study design that the subjects were made aware of their partner's gender at the beginning of the second part of the experiment. Even though no information about their 'new' partner's gender was provided in the third part, gender effects were still stronger than experimentally manipulated status differences.[7] We therefore conjectured that removing the reference to the partner's gender would remove the gender effect completely.

Second, the experimental sessions lasted quite long (more than 30 minutes each) and all subjects were physically present in one room. It is possible that the

[7] Note, however, that if we assume that in low risk games (as explained below), it is greed, not fear, that is the dominating motivation (despite both motivations being present in the game structure), we would expect women to cooperate at higher rates than men (Simpson 2003). If, on the other hand, gender acted as a status characteristic, men (high status) should only cooperate more when they believe they are paired with women (low status), and so we should still observe different rates of cooperation among subjects of unequal status (Sell 1997).

physical proximity and the relatively long time spent together working on the tasks together fostered some group-formation process, whereby subjects developed a sense of common identity and positive feelings for one another, which precluded them from taking actions that might 'hurt,' and instead encouraged them to take actions that could benefit the others, regardless of any information about their respective statuses. Importantly, common identity is conducive to the subjective transformation of a PD into an assurance game (Kollock 1998b), and the transformation is conducive to cooperation (Simpson 2004) because defection is the dominant strategy in the PD but not in the assurance game. It is possible that making the sessions shorter and moving to a bigger computer lab might mitigate or delay the group process.

Third, we suspected that our failure to find any status effect might have resulted from the relatively low stakes in our investment game setting. Recall that each subject was given a *guaranteed payment* of 30 PLN after the first two parts of the study, and that the endowment they received in the third part was worth 5 PLN. The risk associated with 'misplaced trust'– or with allocation of one's whole endowment to their partner – might therefore have seemed rather low to our subjects. Accordingly, for the third part, we decided to decrease the guaranteed payment, and to increase the amount to be invested by the subjects. Also, recall that the endowment received from the experimenter was described in terms of abstract 'tokens,' which might have diminished the subjects' ability to assess the risk appropriately, because such assessment required them to 'translate' the tokens into monetary values in order to estimate the financial effect of a particular 'investment.' Such calculations are quite difficult to process. Consequently, it may have been the case that if we described the endowment in terms of money rather than 'tokens,' subjects would have been more likely to estimate their risks more correctly. With these ideas in mind, we went on to carry out a second study with somewhat modified design.

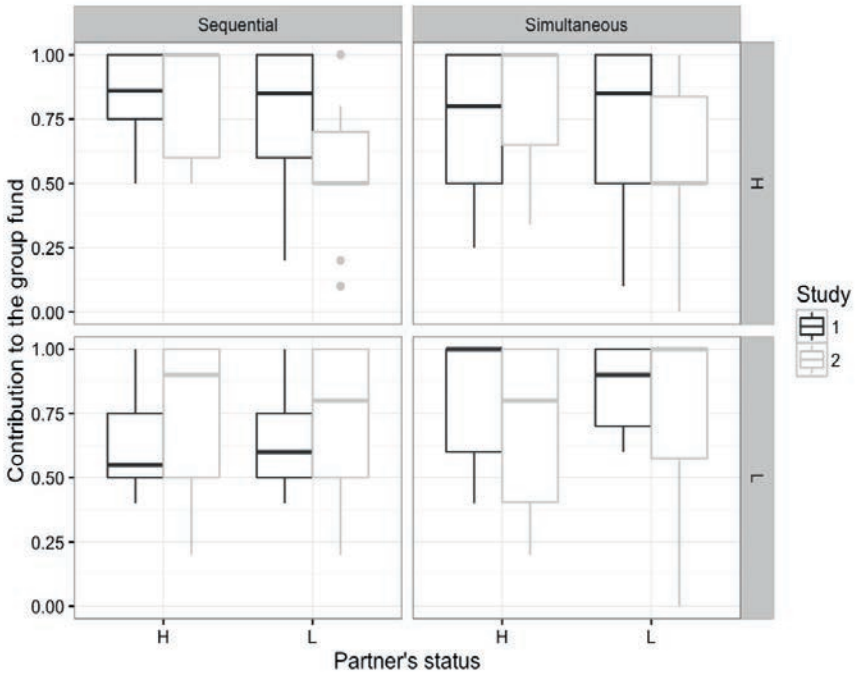
Study 2

The second study was conducted in February and March 2015. As previously, the participants were students from various departments of public and private universities located in Warsaw. We used the same channels of communication to recruit the participants. Once again, the ads and announcements informed students about the possibility of earning money by taking part in a scientific study of decision making processes in groups.

The general design of the second study was the same as in the first, but we changed a number of important details. First, we made the sessions considerably shorter by reducing to 10 the number of tasks in the first and second part of the experiment. Second, we simplified the instructions to make them shorter and clearer. Third, the only information about the 'partner' that we made available to the subjects was his or her contrast sensitivity score – i.e., there was no reference to the partner's age or gender. Fourth, the guaranteed payment after the first two parts of the experiment was reduced to 20 PLN and the endowment given by the

experimenter at the beginning of the third part was increased to 10 PLN. Thus, the guaranteed payment could be doubled by mutual cooperation or remain unchanged in the case of mutual defection. Fifth, the decisions in the investment game were described in terms of monetary stakes and not tokens.

Figure 1: Comparison of the results of Study 1 and Study 2



A total of 135 subjects took part in Study 2, four of whom were excluded due to violating scope conditions. Our final analyses thus include 131 observations. Figure 1, above, shows changes in the results between first and second experiment. In this figure, we make use of box-and-whisker plots to summarize the distribution of the dependent variable across experimental conditions. The rectangular boxes represent quartiles of the distribution, with the bottom of the box corresponding to the first quartile, the top – to the third quartile, and the thick segment inside the box to the median. Further, the horizontal whiskers adjacent to the boxes cover the distance of 1.5 times the inter-quartile range from the median. Any points beyond that range are outliers, represented as dots on the graph.

Figure 1 is divided into four panels and there are four box-and-whisker plots within each panel, corresponding to a total of 16 combinations of the subject's status (high vs. low), the partner's status (high vs. low), game type (simultaneous vs. sequential), and experiment (1 vs. 2). Figure 1 shows some notable differences between the results of the two experiments. The dependent variable in both experiments is the proportion of subjects' initial endowment that they transferred to their partners. The two top panels represent the behaviour of high-status subjects and the two bottom panels represent the behaviour of those who were assigned low

status. Further, the columns of the graph correspond to the type of game, and different colours were used to distinguish the results of our first study from those of the second.

The graph shows substantial differences in contributions across the two experiments, especially in the case of the high-status subjects. In the first experiment, high-status subjects' median cooperation was about the same, regardless of their partner's status or whether the game was simultaneous or sequential (although there is more variation in their contributions in the simultaneous games, as indicated by larger boxes). In the second study, their cooperation clearly varies depending on the partner's status (but not on the type of the game).

Low-status subjects' behaviour also changed from one study to the other, but the change is not as striking as for high-status subjects, and it occurred in sequential games only. In Study 1, low-status subjects' cooperation was rather moderate in the sequential game, regardless of whether their partner was of high or low status. In Study 2, their median cooperation increased, especially in games played against high-status partners.

Let us now return to our hypotheses to see if the data from the second experiment are more consisted with our predictions. First, the median for pairs LH is higher than for pairs HL, which is consistent with the first hypotheses. This result holds for both sequential and simultaneous games, which also confirms our expectations, although it is somewhat less pronounced in the latter compared with the former. Second, as indicated, high-status subjects do vary their behaviour depending on the status of the partner. As Figure 1 makes it clear, their median cooperation rate, in both simultaneous and sequential games, is 100% when their partner is of high status and only 50% if the partner is of low status. However, low-status subjects' behaviour fails to conform to the hypothesized pattern. In sequential games the low-status subjects do cooperate more with high-status partners than with low-status partners, but the difference is much smaller than the corresponding effect for high-status subjects. Also, in simultaneous games, the low-status subjects cooperate more when their partner is low-status than when their partners are high-status, in contrast with our prediction.

On the whole, the data from Study 2 are more consistent with our predictions. We also did not find any significant differences in the rates of cooperation between male and female subjects in the second experiment. This is an important result, because it suggests that gender was not a salient feature of our experimental design, and that is precisely what we anticipated. In addition, the fact that gender had no effect on cooperation in Study 2 bolsters our confidence in the results of the study, as it suggests that we succeeded in removing a potential confounding factor from the picture.

Discussion and conclusions

A new multilevel model explaining the effects of status differences on cooperation in one-shot 2-person PD was outlined in this article. It supports previous theoretical analyses that demonstrate the interrelationship between cooperation

and status phenomena. The results of the empirical test, however, suggest that the theoretical model must be revised and expanded. The observations from the second study are consistent with our hypotheses regarding the behaviour of high status actors. Our analyses show that high-status subjects indeed cooperate more with other high status partners and less with low status partners, which suggests that they form performance (cooperation) expectations based on salient status differences. Also, we did not find any interaction between status and the type of game; this too squares with our hypotheses. Our observations nevertheless are inconsistent with our hypotheses regarding the behaviour of low status subjects, whose decisions seem to be unaffected by information about their partners' status. Whether this means they did not form performance (cooperation) expectations, or that some other status related process (e.g., reward expectations) affected their decisions, requires further theorization.

The results presented above also illustrate the idea that the cumulative growth of knowledge can only be achieved when theoretical models and their empirical replicas are carefully constructed, and when empirical activity is theory-driven. The feedback between model, its experimental replica, and collected data stimulated additional questions and improvements to the experimental design. These improvements had a number of desired effects, including eliminating gender effects.

Note that we modified several features of the experimental design between study 1 and study 2. This precludes us from conclusively identifying the feature that contributed most to the differences in the results between the two experiments. Doing so was beyond the scope of our work, however, as we simply aimed to satisfy the scope conditions of our model, and to identify and remove any feature that might have compromised or confounded the results of the first experiment. Careful investigation into how a particular feature of our study design interacted with our major dependent variable – cooperation in the investment game – is interesting in its own right, as it may shed some light on how the interplay of status and other social mechanisms affect cooperation, but that question is well beyond the scope of our present study.

It was also beyond the scope of the present study to explain the high overall rates of cooperation in both of our experiments. Such high rates are not unusual. In studies of the effects of punishments (e.g., Fehr, Gintis 2007), and shame and honour (Jacquet et al. 2011) reported cooperation was also high. Our average cooperation rate was nevertheless higher than the average of 40-60% in classic PD settings. Again, investigating whether any particular feature of the setting (e.g., status manipulation, subjects solving a set of tasks before the game or simply sharing the same space) had such an effect is interesting in its own right. Our hypotheses did not specify particular cooperation levels, but rather the order of cooperation levels among pairs of interactants (differences between pairs). Therefore, regardless of whether the overall cooperation was high or low, our hypotheses regarding status differences' effects on the behaviours of participants could still be tested and indeed showed such effects for high status subjects.

A theory is never fully complete or finished. Our model can be further elaborated to include joint effects of nominal and status distinctions, as well as status

effects in coordination games and repeated PD. More importantly, however, as our theoretical model links status with expectations of trustworthiness using the concepts of fear and greed, future work on the model may benefit from employing variants of the prisoner's dilemma game in which only one of the two basic motivations are present. More specifically, in one variant of the game the fear component is removed, leaving only the greed component, while in another variant it is the other way around. Given that actors of different status respond differently to fear and greed component of the game, predictions concerning cooperation in the modified games can be developed, leading to a theoretical elaboration of the original model. The design described in the present paper can be thought of as a baseline or standardized setting with which results of the future studies can be compared.

Appendix

The description of the one-shot two-person Prisoner's Dilemma (PD) game[i]

In a PD game, actors choose between "cooperation" (C) and "defection" (D). The actors are assumed to be "rational" in the sense that they are motivated to maximize their own payoff. However, their payoff depends not only on the decision they make, but also on that made by their partner in the game. The payoff structure is conveniently presented in the form of a matrix, such as the one shown below.

Column / Row	Cooperation	Defection
Cooperation	R,R	S,T
Defection	T,S	P,P

There are two players and each of them has two options to choose from, so there are four possible combinations of their decisions, each such combination represented by a cell in the matrix. Each cell contains two numbers – the first corresponds to the payoff of the row player and the second to the payoff of the column player.

The payoffs in our matrix are represented by abstract symbols rather than specific amounts. The symbols are interpreted as *T*: the temptation to free ride on the partner's cooperation, *R*: the reward for mutual cooperation; *P*: the punishment for mutual defection; *S*: the sucker's payoff for unilateral cooperation.

The prisoner's dilemma game is defined by the following two conditions: (a)[ii] the ordering of payoffs is as follows: $T>R>P>S$; and (b)[iii] $2R>T+S$. Any set of numbers satisfying these constraints constitutes the payoff set in the PD.

Let us now consider what rational actors will do in the Prisoner's Dilemma game, beginning with Row's decisions. Suppose that Row expects Column to cooperate. Given this expectation, Row's best response is to defect, because $T>R$. Further, if Row thinks that Column will defect, Row's best response, again, is to defect, because $P>S$. Thus, defection is Row's dominant strategy in the game, as it gives Row a better payoff than cooperation, regardless of what Column does. Because the payoff matrix

is symmetric, the same reasoning applies to Column's decisions. Thus, rational decision makers arrive at (D,D) and each end up having P, which is clearly inferior to the group's optimal outcome (C,C) giving R to each player.

Given the payoff structure, we can define two parameters that represent the significance of two motivations for defection: *fear* and *greed*. Fear is defined as *P–S*, the difference between the payoff for mutual defection and unilateral cooperation, while greed is equal to the difference between the payoff for unilateral defection and mutual cooperation, *T–R*. In most application, the two parameters are equal, but it is possible to vary levels of fear and greed, as long as the defining constraints of the game are satisfied.

Notes added by the Editor

[i] Prisoner's Dilemma is a *two-person game in the normal (matrix) form.* The structure of such a game depends on what *outcomes* are assigned to all action (strategy) pairs; and on how the actors' *preference relations* on the set of all possible outcomes are related to each other.

[ii] This ordering of *payoffs* (outcome utilities) implies that *for Row* the (D,C) outcome, or the outcome produced by Row's unilateral defection, is better than the (C,C) outcome (the result of mutual cooperation). As a consequence, Row may yield to the temptation to defect if he has any reason to suspect that Column is going to cooperate. The (C,C) outcome is preferred in turn by Row to the (D,D) outcome (mutual defection). For Row, the worst (least preferred) of 4 outcomes is (C,D), or the case where Row (the 'sucker') gets exploited by Column. Column's preference relation on the set of outcomes is determined similarly. It is easy to see that the partners' preference relations, on the one hand, are in conflict (the best hope outcome of Row is the worst fear outcome of Column and conversely); on the other hand, the preferences of the players partially agree, as both Row and Column prefer (C,C) to (D,D). Therefore, both players are motivated to cooperate with each other, but at the same time they are tempted to seek each one's own most preferred outcome, which amounts to disregarding the interest of the partner. Prisoner's Dilemma is an example of a *mixed motive game.* The problem with PD is that *individual rationality* prompts to each player the use of *dominant strategy* D to the effect that the game ends up with the outcome (P,P) which is not Pareto optimal (because (R,R) is better than (P,P) for both players). An outcome is called *Pareto optimal* if there is no other outcome which makes at least one player better off without making the other player worse off.

[iii] The inequality (b) makes sense under the assumption that Row and Column use the same *utility scale* to evaluate the outcomes of their co-action. If so, the four *payoffs* T, R, P, S, having the same value for Row and Column, can be operationally defined in terms of a number of units of a certain resource that is valued similarly by both players. The numbers 2R=R+R, T+S, and 2P are then regarded as the amounts of that resource possible to be earned by the 2-person group. Notice that (b) and the inequality 2R>2P, which follows from (a), jointly imply that 2R is the *maximum total group payoff.* Under such an interpretation the PD game can also serve as a mathematical model for any social interaction system in which the actors, apart from pursuing their *individual goals*, may be *collectively oriented*, in other words, they may show concern for the benefit of the group as a whole.

References

Ahn T.K., Ostrom E., Schmidt D., Shupp R., Walker J. (2001). *Cooperation in PD Games: Fear, Greed, and History of Play*. Public Choice, 106, p. 137–155.

Aksoy O. (2015). *Effects of Heterogeneity and Homophily on Cooperation*. Social Psychology Quarterly, 78, p. 324–344.

Axelrod R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Berg J., Dickhaut J., McCabe K. (1995). *Trust, Reciprocity, and Social History*. Games and Economic Behaviour, 10, p. 122–142.

Berger J. (2007). *The Standardized Experimental Situation in Expectation States Research: Notes on History, Uses, and Special Features*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, San Diego: Academic Press, p. 353–377.

Berger J., Wagner D.G., Zelditch M., Jr. (1987). T*heory Growth, Social Processes, and Metatheory.* Working Paper no. 1987-2. Department of Sociology, Stanford University.

Berger J., Webster M., Jr. (2006). Expectation, Status, and Behavior. In: P.J Burke (ed.), Contemporary Social Psychological Theories. Stanford: Stanford University Press, p. 268–300.

Berger J., Zelditch M., Jr. (1997). *Theoretical Research Programs: a Reformulation*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes.* Stanford, CA: Stanford University Press, p. 29–46.

Brewer M.B. (1979). I*n-group Bias in the Minimal Intergroup Situation: a Cognitive-Motivational Analysis.* Psychological Bulletin, 86, p. 307–324.

Brewer M.B., Kramer R.M. (1986). *Choice Behavior in Social Dilemmas: Effects of Social Identity, Group Size, and Decision Framing*. Journal of Personality and Social Psychology, 50, p. 543–549.

Cohen B.P. (1989). *Developing Sociological Knowledge: Theory and Method*, 2nd ed. New York: Nelson-Hall.

Cohen, B.P. (1980). *The conditional nature of scientific knowledge*. In: L. Freese (ed.), *Theoretical Methods in Sociology*. Pittsburgh, PA: University of Pittsburgh Press, p. 71–110.

Dawes R.M., Orbell J.M., Simmons R.T., Van De Kragt A.J.C. (1986). *Organizing Groups for Collective Action*. American Political Science Review, 80, p. 1171–1185.

Dawes R.M., Van De Kragt A.J.C, Orbell J.M. (1988). *Not Me or Thee but We: The Importance of Group Identity in Eliciting Cooperation in Dilemma Situations: Experimental Manipulations*. Acta Psychologica, 68, p. 83–97.

Fararo T., Skvoretz J. (1993). *Methods and Problems of Theoretical Integration and the Principle of Adaptively Rational Action*. In: J. Berger, M. Zelditch (eds.), *Theoretical Research Programs. Studies in the Growth of Theory.* Stanford, CA: Stanford University Press, p. 416–450.

Fehr E., Gintis H. (2007). *Human Motivation and Social Cooperation: Experimental and Analytical Foundations.* Annual Review of Sociology, 33, p. 43–64.

Foddy M., Yamagishi T. (2009). *Group-Based Trust*. In: K.S. Cook, M. Levi, R. Hardin (eds.), *Whom Can We Trust.* New York: Russel Sage Foundation, p. 14–41.

Foschi M. (1997). *On Scope Conditions*. Small Group Research, 28, p. 535–555.

Habyarimana J., Macartan H,, Posner D.N., Weinstein J.M. (2001). *Coethnicity. Diversity and the Dilemmas of Collective Action.* New York: Russell Sage Foundation.

Hogg M. (1996). *Intragroup Processes, Group Structure and Social Identity*. In: W.P. Robinson (ed.), *Social Groups and Identities.* Oxford: Butterworth-Heinemann, p. 65–93.

Jacquet J., Hauert C., Traulsen A., Milinski M. (2011). *Shame and Honour Drive Cooperation*. Biological Letters, 7, p. 899–901.

Karpiński Z., Skvoretz J. (2015). *Repulsed by the 'Other': Integrating Theory with Method in the Study of Intergroup Association*. Sociological Theory, 33, p. 20–43.

Kemper T.D. (1991). *Predicting Emotions from Social Relations.* Social Psychology Quarterly, 54, p. 330–342.

Kemper T.D. (1987). *How Many Emotions Are There? Wedding the Social and Autonomic Structure*. American Journal of Sociology, 93, p. 263–289.

Kollock P. (1998a). *Social Dilemmas: The Anatomy of Cooperation*. Annual Review of Sociology, 24, p. 183–214.

Kollock P. (1998b). *Transforming Social Dilemmas: Group Identity and Cooperation*. In: P.E. Danielson (ed.), *Modelling Rational and Moral Agents.* Oxford: Oxford University Press, p. 186–210.

Kramer R.M., Brewer M.B. (1984). *Effects of Group Identity on Resource Use in a Simulated Commons Dilemma.* Journal of Personality and Social Psychology, 46, p. 1044–1057.

Kumru C.S., Vesterlund L. (2010). *The Effect of Status on Charitable Giving*. Journal of Public Economic Theory, 12, p. 709–735.

Lawler E.J. Ridgeway C.L., Markovsky B. (1993). *Structural Social Psychology and the Micro-Macro Problem*. Sociological Theory, 11, p. 268–290.

List J.A. (2007). *On the Interpretation of Giving in Dictator Games*. Journal of Political Economy, 115, p. 482–493.

List J.A., Cherry T.L. (2008). *Examining the Role of Fairness in High Stakes Allocation Decisions*. Journal of Economic Behavior & Organization, 65, p. 1–8.

Lovaglia M. (1997). *Status, Emotions, and Structural Power*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes.* Stanford, CA: Stanford University Press, p. 159–178.

Lovaglia M.J., Houser J.A. (1996). *Emotional Reactions and Status in Groups*. American Sociological Review, 61, p. 867–883.

Lovaglia M., Willer R.B. (2002). *Theory, Simulation, and Research: The New Synthesis*. In: J. Szmatka, M.J Lovaglia, K. Wysieńska (eds.), *The Growth of Social Knowledge*, Westport, CT: Praeger, p. 247–262.

Markovsky, Barry. 1994. *The Structure of Theories*. In: M. Foschi, E. Lawler (eds.), *Group Processes: Sociological Analyses*. Chicago, Ill: Nelson-Hall, p. 3–24.

Markovsky B. (1997). *Building and Testing Multilevel Theories*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes.* Stanford, CA: Stanford University Press, p. 13–28.

Molm L.D., Takahashi N., Petersen G. (2000). *Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition*. American Journal of Sociology 105, p. 1396–1427.

Rapoport A., Chammah A.M. (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor: University of Michigan Press.

Ridgeway C.L. (2014). *Why Status Matters for Inequality*. American Sociological Review, 79, p. 1–16.

Ridgeway C.L. (2006). *Status Construction Theory*. In: P.J. Burke (ed.), *Contemporary Social Psychological Theories.* Stanford: Stanford University Press, p. 301–323.

Ridgeway C.L. (2001). *Inequality, Status, and the Construction of Status Beliefs*. In: J.H. Turner (ed.), *Handbook of Sociological Theory*. New York: Springer, p. 323–340.

Ridgeway C.L. (1982). *Status in Groups: The Importance of Motivation.* American Sociological Review, 47, p. 76–88.

Sell J. (2007). *Social Dilemma Experiments in Sociology, Psychology, Political Science, and Economics*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences.* San Diego: Academic Press, p. 459–481.

Sell J. (1997). *Gender, Strategies, and Contributions to Public Goods.* Social Psychology Quarterly, 60, p. 252–265.

Shelly R.K. (2002). *How Scope and Initial Conditions Determine the Growth of Theory*. In: J. Szmatka, M.J. Lovaglia, K. Wysieńska (eds.), *The Growth of Social Knowledge*, Westport, CT: Praeger, p. 115–130.

Simpson B. (2006). *Social Identity and Cooperation in Social Dilemmas*. Rationality and Society, 18, p. 443–470.

Simpson B. (2004). *Social Values, Subjective Transformations, and Cooperation in Social Dilemmas*. Social Psychology Quarterly, 67, p. 385–395.

Simpson B. (2003). *Sex, Fear, and Greed: A Social Dilemma Analysis of Gender and Cooperation*. Social Forces, 82, p. 35–52.

Simpson B., Willer R., Ridgeway C.L. (2012). *Status Hierarchies and the Organization of Collective Action*. Sociological Theory, 30, p. 149–166.

Skvoretz J. (1998). *Theoretical Models: Sociology's Missing Links.* In: A. Sica (ed.), *What Is Social Theory? The Philosophical Debates*. Malden, MA: Blackwell Publishers, p. 238–252.

Tajfel H. (1982). *Social Psychology of Intergroup Relations*. Annual Review of Psychology, 33, p. 1–39.

Troyer L. (2002). S*ES V7.2.* A Computerized Version of the Expectation States Research Program's Standardized Experimental Setting.

Turner J.C. (1985). *Social Categorization and the Self-Concept: A Social Cognitive Theory of Group Behavior*. Advances in Group Processes, 2, p. 77–121.

Wagner D. (2000). *On the Irrationality of Rejecting Falsified Theories*. Sociological Focus, 32, p. 27–39.

Walker H.A., Cohen B.P. (1985). *Scope Statements: Imperatives for Evaluating Theory*. American Sociological Review, 50, p. 288–301.

Willer R. (2009). *Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem*. American Sociological Review, 74, p. 23–43.

Wysieńska K., Szmatka J. (2002). *Positivism and Theory Construction in Group Processes*. In: J. Szmatka, M.J. Lovaglia, K. Wysienska (eds.), *The Growth of Social Knowledge.* Westport, CT: Praeger, p. 77–96.

Yamagishi T., Yamagishi M. (1994). *Trust and Commitment in the United States and Japan*. Motivation and Emotion, 18, p. 129–166.

Yamagishi T., Kiyonari T. (2000). *The Group as the Container of Generalized Reciprocity.* Social Psychology Quarterly, 63, p. 116–132.

Yamagishi T., Mifune N. (2008). *Does Shared Group Membership Promote Altruism? Fear, Greed, and Reputation*. Rationality and Society, 20, p. 5–30.

Yamagishi T., Jin N., Kiyonari T. (1999). *Bounded Generalised Reciprocity: Ingroup Boasting and Ingroup Favoritism*. Advances in Group Processes, 16, p. 161–197.

Yamagishi T., Makimura Y., Foddy M., Matsuda M., Kiyonari T., Platow M.J. (2005). *Comparison of Japanese and Australians on Group Based Cooperation.* Asian Journal of Social Psychology, 8, p. 173–190.

Modelowanie sytuacji społecznych: zaufanie i kooperacja pomiędzy nieznajomymi o nierównym statusie

Niniejszy artykuł opisuje proces konstruowania i testowania nowego modelu teoretycznego wyjaśniającego, jak nierówności pod względem statusu wpływają na poziom kooperacji między aktorami uczestniczącymi w jednokrotnie rozgrywanym Dylemacie Więźnia (DW). Łącząc idee pochodzące z teorii charakterystyk statusu i działań zbiorowych, przewidywaliśmy, że w DW, w którym aktorzy są widocznie zróżnicowani pod względem statusu, poziom kooperacji będzie zależał od statusu aktora w relacji do statusu jego partnera, a także tego, czy gra ma charakter symultaniczny czy sekwencyjny. Aby zademonstrować wyzwania związane z konstruowaniem i testowaniem nowych modeli teoretycznych, przedstawiamy dwa różne skonstruowane przez nas eksperymenty dotyczące dwuosobowych jednokrotnie rozgrywanych sytuacji DW. Artykuł wieńczy dyskusja o charakterze substantywnym i metodologicznym.

Słowa kluczowe: status, spolegliwość, Dylemat Więźnia, budowa teorii