HARRIS SCHOOL WORKING PAPER
SERIES 06.05J

CHAPTER 8: NSNX EFFECTS IN THE PUBLIC GOODS GAME *

Howard Margolis

**This is a draft chapter from* Cognition and Extended Rational Choice *(forthcoming Routledge 2007).*

Chapter 8. NSNX effects in the Public Goods game

Without stark demonstrations, no one would suppose that *visual* misperceptions like the Muller-Lyer illusion in figure 1 could occur. Similarly, gross *logical* illusions parallel to the surprising but indisputable visual illusions would not seem plausible without examples strong enough to catch the person who needs to be convinced. But we now have examples at hand where sophisticated subjects (like people who might read a book like this) are prompted to clear intuitions that make no sense, as all of us see an unmistakable apparent difference in length that does not exist in the Muller-Lyer drawing. And without the evidence of experiments like those reviewed in chapter 7, defaulting to inappropriate *social* contexts would seem implausible, though here it is harder to convince a reader that she too can be caught. But I will make an attempt in Chapter 10. But we have enough in hand (from Chapter 7) to leave no reasonable doubt that students in experiments can be strikingly caught, which in fact makes it rather naïve to doubt that professors can be caught too.

FIG 1 HERE

Sometimes illusory effects must be only an artifact of odd conditions peculiar to an experiment. But there is no reason to suppose that the effects only occur in a laboratory, since it is certainly not only in a laboratory that people sometimes face contexts that are unfamiliar and lacking the dense and usually redundant cues of natural situations. So as I have urged more than once, the possibility that adverse defaulting might play a perverse role in real situations warrants attention. To get to that we need to have in hand some experience in applying NSNX reasoning to data. But in Chapter 7 I specifically avoided experiments which could provide an occasion for that, since what we wanted were results which were not contingent on NSNX or any other form of other-regarding motivation . So examples were selected for the special property that purely self-interested and purely group-interested motivation would yield the same results from a player responding sensibly to the incentives she faces.

But it was easy to find experiments in which players did not respond sensibly at all. And exploring what might account for that in Chapter 7 allows us now turn to data that might profit for a NSNX analysis, and which indeed requires some allowance for other-regarding motivation to make any reasonable sense. But we do that armed with an account (from Chapters 6 and 7) of how players might be vulnerable to distortions of their own preferences. As I write, efforts to account for positive contributions in cooperation games in terms of error-prone but essentially strictly self-interested motivation still continue. I don't think I distort the situation to treat this as a fading enterprise. An account solely in terms of NSNX, however, also is unlikely to be sufficient, given the strong cognitive effects we have been seeing. Therefore we need an analysis in terms of NSNX + cognition, since an account in terms of NSNX alone (or in terms of any other model of motivation alone) that is missing a cognitive component may cripple the attempt to gain insight from data.

An exchange between two Nobel laureates (Paul Samuelson and Milton Friedman) is relevant here. Friedman had made his argument for boldly unrealistic assumptions, using the example of ignoring resistance of the medium in analyzing free fall. This indeed is how Galileo proceeded, with memorable results. But, Samuelson (1963) objected, what if the object you are analyzing is a parachute? Plainly, it is nice to keep things simple. But two simple things at work can yield what is hopelessly complex unless you allow for both components.

I start from some fresh analysis of two particularly well-known sets of Public Goods data, and then consider a recent (and still unpublished as I write, hence citations to working papers) but particularly suitable set of variants on the standard game./1 The cognitive component of the discussion becomes more important as we proceed, but never becomes as absolutely central in this chapter as it will in the pair of chapters to follow. But there is a cognitive aspect in play throughout.

Start with a few general remarks, some in the nature of review of points made earlier but specifically needed here. In the basic Public Goods game, all players share equally in a pool of contributions. The return to each player per token in the pool is less than 1 but the sum shared among all players is greater than 1. So it goes against the self-interest of a player to contribute (since the return is < 1, there is a free-rider incentive), but it is profitable to contribute as part of a mutual commitment of all to contribute (the sum of the individual returns is > 1). But mutual commitment is not available. The experiment guarantees anonymity, and strictly forbids communication or any other means by which commitments might be made, intentions shared, punishments or rewards implemented. These features were intended to maximize the prospect that (at least by the end of a series of rounds), cooperation would be nil. This would provide a baseline from which to investigate design of institutions which could yield successful cooperation. But it turned out that while contributions typically decline over the series of rounds, the baseline could rarely be reached.

Characteristic results of this basic game have proved to be highly consistent across experiments. Nationality, gender, and various subtler distinctions have made little difference. (Still the most widely-cited review is Ledyard 1995.) But as mentioned in Chapter 7, across the entire range of subject pools, players in every culture come to the game with experience with common pool problems. They are a pervasive aspect of social life, encountered in some form every day at home and work and even in chance encounters with strangers. We can expect players to recognize that the context of a common pool game is on the cooperative branch of the cascade introduced in chapter 7. And then we can also expect subjects to ordinarily respond to the risky cooperation frame (2a) that correctly matches the context the game puts them in. But we saw puzzling choices in Chapter 7 that turned on the difficulty of escaping this default when the game is altered to make that inappropriate -- and in either direction. We saw degenerate games in which a cooperative frame (even the weakly cooperative 2a) is a mistake, and we also saw degenerate games in which make failing to move from the weak cooperation to the strong cooperation frame (2b) is a mistake. In both directions players seemed vulnerable to being inappropriately caught by the default. We see choices that the chooser herself would find puzzling.

In a standard (not degenerate) Public Goods game, the "ordinarily" qualifier in the previous paragraph is needed to allow that some subjects of a particularly cooperative bent might escape the risky competition default (frame 2a in the cascade) for the coordination frame (2b). They might (initially, for they are almost certain to find things otherwise) take it for granted that fellow players will also see the game that way. At the other end, a player well-drilled in the standard economic model of rational self-interest, and since this is after all a game and games are usually played competitively, might be on the competitive branch of the cascade though with cues adequate to escape from the zero-sum frame (1a) to the payoff-maximizing frame (1b). Neither of these responses involves adverse defaulting in the invidious sense developed in Chapters 6 and 7. A bright player could find herself seeing the game in either of these ways and defend her view of the game as reasonable. But since typical results are much the same across experiments and across subject pools, propensities to both departures are apparently not very different across subject pools. We ought to find NSNX effects that generalize across subject pools.

The *updown* effect

As in all standard Public Goods games, parameters in the experiments of the first half of this chapter were chosen such that the group does best if everyone contributes their entire endowment, but each individual does best for herself if others contribute while she keeps her endowment. Let *give* = a player's choice in the current round, *pool* = sum of all *gives* in this round, *y* = endowment per round, *n* = number of players in each group, and *a* = return to each player for each token in the common pool. A player's total payoff for the round will be *y* - *give* + *a*pool*. The cost of contributing a token is 1 - *a* (a fraction of each token given comes back to the donor). The gain to the group from any player contributing a token is *n*a* - 1, since the each member of the group (including the donor) gets the return from any member's contribution of a token.

The value ratio (G'/S') on the right in the NSNX equilibrium condition would then be subject to various complications noticed in Chapter 1, but that ratio would be anchored on (*n***a*-1)/(1-*a*). So other things equal, the value ratio increases with increases of *n* or *a*. But W on the left of the equilibrium condition (W = G'/S') responds to how a player's contributions look relative to what others are contributing. To catch that aspect of a Public Goods game, define *gavg* as the *give* of a player in a round relative to the average *give* of other players. So *gavg* = *give* / ((*pool*-*give*)/(*n*-1)).

Further interpretation is tied to the Darwinian argument that underpins NSNX in Chapter 1. Darwinian variation is blind and selection near-sighted, which creates a challenge for any account that postulates choice motivated by group- as well as self-interest. The problem is not insurmountable. In SA&R (Chapter 3), I suggested that an evolution of gene linkages across traits that could favor selection "as if" a longer view were effective. But the main point of the Darwinian argument, sketched in Chapter 1 here, is that we can see that human beings exhibit a degree of other-regarding as well as self-interest motivation, and thinking in Darwinian terms about how that might be sustained leads us to the pair of NSNX rules, which in turn imply the NSNX equilibrium condition.

But concern for Darwinian plausibility warns us not to treat the G'/S' ratios of the NSNX model as more than ordinal. A creature might get by with only a sense of greater-than or lesser-than to manage choice between alternatives. Nicely-refined quantitative assessments would be better (if there were no costs to managing that and a feasible evolution pathway were available), but it does not look like we have been provided with that until cultural rather than genetic evolution yielded such things as literacy and recording devices. Our deeply entrenched propensities long antedate that.

So the Darwinian view allows for and indeed at least weakly predicts the predominance of qualitative over quantitative effects we have seen repeatedly, yielding the sometimes startling insensitivity we have seen of intuitive judgments to even quite gross shifts in parameters. Several examples in Chapter 7 exhibited the stickiness of neglect defaults in the face of starkly contrary quantitative cues. And as mentioned there, parallel insensitivity to quantitative parameters have been noticed in many other studies entirely unconnected with data from cooperation experiments. The effect of quantitative cues is likely to be weaker than logic would warrant. And the value ratio is several steps removed from the cues. A general consequence is that we should not expect more than comparative statics inferences from the NSNX argument. But as will be seen that is sufficient to yield strong results.

Under strictly self-interested choice, players would contribute nothing to a Public Goods game, on a variant of the backward induction argument in the 100 PD game noticed at the end of chapter 1. But NSNX players should ordinarily be willing to contribute something to a common pool (even when there is no way to punish a free-rider), with the "ordinarily" qualification allowing for the various atypical cases mentioned earlier. The positive contributions expected under NSNX, however, should show the effects of "neither selfish nor exploited" mixed motivation, which would be adversely affected by the difficulty in coordinating contributions which the standard conditions (anonymity, etc.) are in fact designed to impose.

An easy comparative statics inference from Rule 1 (NSNX efficiency), however, is that even without explicit coordination, contributions should tend to increase with increasing return per token from the pool and also with increasing membership in the group that will benefit. On the standard account players should not contribute at all. But increasing cooperation with increasing return is easy to accommodate even in terms of the standard account. *Given* that (for whatever reason) players contribute at all, it would be surprising if they did not tend to contribute more if the per token cost of contributing (1 - *a*) falls. And this is a well-marked feature of Public Goods data. On the other hand, a tendency for contributions to also increase with increasing *n* is puzzling on the standard account (Ledyard 1995), since a player concerned only with self-interest tautologically does not value a better return to the group, only lower cost to self. On that logic, increasing *a* motivates contributing but increasing *n* does not, even allowing that somehow players do contribute. But contributions do generally increase with increasing *n* as well as with increasing *a*, as NSNX would expect. Since the NSNX argument was published in 1982, generally increasing cooperation with increasing *n* as well as for increasing *a* is confirmation of an implication antedating the observations.

On the prevailing view, some fraction of players can be labeled "conditional cooperators", and attempts to sort that out typically find the fraction to be large. And since an immediate consequence of the NSNX equity rule (p.) is that *in general* players must be conditional cooperators, that observation, in terms of NSNX, must hold. But NSNX implies that efforts to "type" players as conditional cooperators or not conditional cooperators should lead to only superficial results. "Types" (other than conditional cooperators) should usually be unstable, so that it does not take much to reveal that they are conditional cooperators after all. I will have occasion to provide several examples as we proceed. Even the player who sees the game competitively will ordinarily be a NSNX player who is seeing free-riding as akin to bluffing in poker… just part of the game, not a strictly self-interested player who recognizes the game as cooperative but is indifferent to norms of reciprocity.

Unlike a strictly self-interested player, a NSNX non-cooperator's frame might shift as cues to what kind of game he is in accumulate during play. On the other hand, a cautious player who in fact would want to contribute all his tokens if he knew others would, might start by contributing zero, waiting to see what others do. At another extreme a bold player, but one seeing self-interested choice as a hard-nosed but not unfair way to play, might still contribute some chips in early rounds to encourage others to give enough in later rounds to make that profitable. And another player might give his entire endowment, or give nothing, but for no better reason than that he misunderstood the instructions or let his attention wander.

So as every experimenter learns, no inference can be drawn from the behavior of one subject, and especially one subject's choice in one round. But strong effects can show through these atypical responses. If NSNX is right, players are ordinarily trying for a "neither selfish nor exploited" balance, and overall results should reflect that even through the noise in the data introduced by complications of the sort I've just sketched.

The most obvious such inference is that on the "neither selfish nor exploited" argument, a player who has given more than the average in her group ought to be more inclined to give less in the next round than a person who has given less than the average. Other things equal, when a player sees she has given more than others, then directly from Rule 2, W increases. But, other things equal, G' and S' are the same. The see-saw on p. – would, consequently, tilt toward more weight to self-interest. The player's sense of how she stands with respect to a "neither selfish nor exploited" equilibrium shifts toward the exploited side. And we should see a converse tendency if she has given less than others in her group. Hence if *gavg* is high (player's *give* is big compared to what others in the group have given) the propensity to give in the next round should fall. And more generally, if motivation is NSNX the prospects of cooperation in future rounds must vary with whether some players are contributing a lot and making minimal gains while others are contributing nothing and gaining a lot… as against the case where everyone is contributing and gaining about as much.

Players are not ordinarily given information about the distribution of contributions, and in the data used here they were not. The average given by others is implicit in a player's own payoff for the round, but it is not transparent, which would reduce its impact in the unfamiliar context of an experiment, as repeatedly seen in the odd cases of Chapter 7. And even when detailed information on giving by others is available, the weakness of quantitative inputs relative to qualitative impressions will diminish its impact. Consequently, in addition to what a player might surmise about how her contribution compares to what others have contributed, the propensity to contribute less (or contribute more) in the next round would also be influenced also by a comparative *give/get* effect, which more directly confronts the chooser and requires only a qualitative perception.

A player who gave more in the previous round but got less can hardly fail to notice that disagreeable news, and must be moved towards feeling exploited, hence be more likely to give less in the current round than to give more. A player who gave less but got more should tend to give more in the next round. If he changes his contribution, it should be more likely to change up than to change down.

But the "gave-more/got-less" influence should be stronger than the "gave-less/got-more" influence, since for "gave-more/got-less" there is a double effect favoring giving less in the next round but for "gave-less/got-more" two opposing effects. *Ex ante*, a player has to rely on some intuition about G'. He knows S' (it is coming out of his pocket), but until he gets feedback after the round, he does not know the value of choosing as he has. If he were to turn out to be the only cooperator, G' could hardly be anything but negative. There is no social value in rewarding a bunch of free-riders, and all the more so when the sole contributor ends up with a loss, while free-riders profit. On the NSNX norms logic of Chapter 1, that is not likely to be seen as a social gain at all. In a natural situation, details might be such that that something different is seen. The sole contributor might be reciprocating prior behavior or expecting reciprocity tomorrow, or be far better off than others, or in various other ways be quite willing to be the only contributor. But in the artificial context of a Public Goods game marked by anonymity and no communication, there is not much opportunity for that.

Ordinarily the player who gave more but got less should feel exploited, and in addition in simple stimulus/response terms he was punished for that choice (got less). Call this the *exploited & punished* case. Among *exploited & punished* players who change their contribution in the next round, what I will call *fractiondown* should be high. For the converse "gave-less/got-more" *selfish & rewarded* case, *fractiondown* ought to be lower (a smaller fraction of changes should be to give less). But that must be a weaker effect since here the NSNX effect and the stimulus/response effect go in opposite directions.

Table 1 tabulates (p. 11) these effects for the “high gavg, gave more, got less” case, using data from the two widely-cited series of experiments mentioned earlier, which happened to be among the to be among the earliest sent to me. The results are very similar for other datasets. The main effects prove to be both large and robust, not only in the 13 experiments reported in Table 1, but in many others I have had an opportunity to test. An interested reader can use the “template” described in the Appendix to run this and other tests on data of her own choice . The difference (*exploited & punished* vs. *selfish & rewarded*) is exhibited for all 9 experimental conditions reported in Isaac, Williams and Walker 1994 and for all 4 conditions reported in Fehr & Gachter 2000.

FIG. 2 HERE (table 1)

Could these strong results reflect merely regression to the mean? Table 2 reports (p. 12) summary results that control for level of contribution. The stratified data cover the great majority of non-zero choices (all non-zero *gives* divisible by 5), arranged into High, Mid and Low with respect to "*gavg*"./2 The fourth column for each contribution level sums the results for that level. If the results in Table 1 reflected merely random regression-to-the-mean, we would see no effect looking across the High, Mid, Low columns within contribution levels. Instead we should see a big effect looking at the summed results across levels. To the extent the results are pure NSNX effects we should see the opposite. And what you can see scanning across contribution levels in Table 2 is that there is no sign of regression to the mean, but ample evidence of the NSNX tendency of *fractiondown* to respond to *gavg* in the predicted way. The only clear anomalies are in cells where data points are so sparse that occasional anomalies would be expected.

FIG. 3 HERE (table 2)

"NSD" effects.

If others are contributing in a cooperative game, the salient choice for a NSNX chooser will be to contribute about as much as others, as in the everyday situation of restaurant tipping, people mostly want to do about what others are doing. But how could the player be confident that others are seeing it the same way and playing the same way? On everyday intuition as well on the NSNX-constrained norms sketched in Chapter 1, it would in fact be socially perverse to allow exploitation unless that was sufficiently (meaning, more than just barely) offset by the gains to cooperators. Consequently the social value of a contribution (G') will be influenced not only by the return to the group it will bring in the current round but also by what giving in prior rounds suggests about the prospect of sustaining fair cooperation in future rounds.

Define *nsd1* as the normalized standard deviation of contributions in round 1 (the standard deviation among *gives* relative to the mean *give*). High mean give favors lows *nsd1* (the denominator is large), but so does more or less equal sharing of effort even if the mean is low (the numerator is small). On the argument, prospects for sustaining cooperation should be affected by how coordinated the group happens to be in round 1. Low *nsd1* should enhance and a high *nsd1* diminish G' adjusted for this effect of the prospect of fair cooperation in future rounds. So from the NSNX perspective, a group that fortuitously starts off well-coordinated should have a better chance to stay in step than a group that starts off badly coordinated. In a "before seeing data" note soliciting datasets for this project, I suggested that if NSNX is right, giving to the pool averaged over all later rounds should correlate with what I am here calling *nsd1*. Call this a *firstgrand* effect.

In contrast to the *updown* effect, this *firstgrand* effect should be seen only in "partners" experiments, where group membership stays fixed across rounds, and not in a "strangers" experiment where players are shuffled after each round, allowing no chance for a coherent tendency across rounds./3 The effect would be clearest at the extremes, reflecting the dichotomizing ("twoness") tendency seen elsewhere. A log regression would fit such data better than a linear regression. The charts to follow (p. 14) consequently plot log regressions for giving over all remaining rounds (*grandmean>1*) against the fortuitous coordination in round 1 measured by *nsd1*./4

Figures 2a, 2b, and 2c plot log regressions for data from the most relevant experiments from the series used for the *updown* tests (IWW1994 *n* = 4 and *n* = 40 trials with *a* = .3, plus the FG2000 partners trial -- FG5 -- which was not conditioned by an immediately preceding series of rounds with punishment./5 The predicted effect is evident, but with a cognitive nuance. The data divides into contrasting modest payoff cases (here IWW4.3 and FG5) and high payoff cases (here IWW40.3). For the modest payoff cases, a player who gave well above the mean of what others in the group gave would be left a payoff below her original endowment./6 A player in this situation can hardly fail to notice she is being exploited. But for the high payoff case (with n = 40, a = .3) each token in the pool yields 12 tokens to be shared, so that even if there were many complete free-riders – it is an interesting point to be taken up in a moment that in fact there were very few complete free-riders -- a generous contributor still comes away with a profit. A player can do very well even if grossly exploited relative to free-riders in her group.

FIGS 4a, b, c & d shld be printed as a set HERE

Hence we might expect that the relative importance of standard deviation and mean in round 1 would be different for the two cases. And as the figures show, there is indeed a strong *firstgrand* effect, but separately plotting *grandmean>1* against *standard deviation(rd*

1) and against *mean(rd 1)* suggests the effect is dominated by one or the other component. We see either a response mainly to *standard deviation(rd 1)* or a response to *mean(rd 1)*, rather than a simple response to the ratio. Again, a reader can examine this using the "template" described in the Appendix, which includes the data for these and many other experiments and which can easily take in new datasets of a reader's choice. The dichotomous effect is governed in the way just suggested. In modest payoff experiments (where exploitation is easily severe) the response is mainly to *standard deviation(rd 1)*, and the converse in the high payoff experiments. A related and perhaps more striking feature will be seen in the *lastround* effect coming later.

Contrary to what a first glance might suggest, the effect for the IWW 40.3 (figure 2c) case is not very different from that for the n = 4 cases. The range of variation is much narrower in IWW40.3 than in IWW4.3. Within the narrower interval, the effects are very similar despite the difference in the dominant aspect just discussed.

Although these experiments have been studied for several decades, the relations apparent in these figures have not been noticed, most likely because almost all experiments are in the modest payoff category. But the effect of the mean is weak there; it is the effect of the standard deviation in round 1 that is strikingly large. It is a significant point in favor of NSNX, I think, that what had not been noticed over several decades of discussion of public goods experiments was found almost immediately here, since the theory quite directly demands what turns out to be a fruitful conjecture.

Is the marked effect in Figure 2a only an artifact due to an outlier? The very similar effect in Figure 2b suggests it is not, as does the theory. But as with the updown effects, an interested reader can use the template to check such effects in data of their own choosing.

Finally, Figure 2d shows a stronger-looking but harder to interpret result. This example uses the IWW 40.3 experiment, plotting the *grandmean>1* for each of the 12 groups as a function of the average normalized standard deviation across all of the first 9 rounds of these 10-round experiments (in contrast to Fig. 2c, where the independent variable is *nsd* for round 1 only.)

In Fig. 2d, the round 10 *nsd* is excluded from the across-rounds average since there are no future rounds it could influence, and *mean(rd 1)* is excluded from the grand mean since there is no prior round *nsd* to influence it. The correlation is very tight indeed, but since in eight of the ten rounds (rounds 2-9) a term of the denominator of the independent variable is also a term of the dependent variable, some inverse slope could hardly be avoided. Trials with random data, consequently, also show tight fits on this measure. But for actual data the variance across groups, the slope of the regression, and the r-squared values are all about twice as large as for random data. And *strangers* data, where partners are re-shuffled from round to round, show no overall *nsd* effect. But since random data show a spurious correlation, why should there be no such effect with real data from a strangers experiment? Perhaps that is somehow due to player responses to the absence of expected responses by their "partners", even though they have been told there are no partners. So this result is puzzling. Nevertheless, the contrast between real data and pseudo-data is sufficiently strong that *something* of interest is apparently going on here.

A lastround effect

The groups in all the IWW trials were drawn by random assignment of players from a common pool of students in introductory economics courses. With an interesting exception to be noted, payoffs were in extra course credits not money, but calibration with trials showed no systematic difference. But the IWW results for groups of 40 with a return per player of .3 look very different from the results for groups of 4 with the same return, as should be expected if motivation is NSNX. Direct social value of giving (G' with no adjustment for prospective cooperation) is ten times as large when n = 40. As already mentioned, for IWW40.3, each token given to the pool produces 40 × 3 = 12 tokens for distribution across the group, while IWW4.3 yields only 4 × .3 = 1.2 tokens per token given. With *n* = 4, it takes 100% cooperation to make a social gain unless gain to the free-rider is treated as a social gain rather than as socially perverse. And in terms of contributors self-interest, it is emphatically perverse: they are left worse off than if they had not played at all. If there were a free-rider and three contributors in a round, each giving some common amount, then the three contributors would all be worse off than if everyone gave nothing.

But with *n* = 40 contributors of some common amount would gain even if 90% of the remaining players (36 of 40) were free-riders. These contributors might very reasonably see themselves as exploited by the free-riders, but they would not be outright losers. So we might expect, and certainly NSNX would expect, a difference in behavior between the *n* = 4 and *n* =

40 cases. And there is indeed a large difference. The point requires serious care in drawing any generalizations about cooperation outside the lab from the far more common small *n* experiments. For although the IWW large group trials are atypical relative to other experiments (where *n* = 4 is the most common), it is large group results (not small group results) that are more typical of empirical situations in which free-rider difficulties become critical. Outside the lab, public goods contexts where cooperation is problematical involve numbers that are large not small, and the value of the public good provided by all contributions is often vastly larger for each individual than the cost to individual of his own contribution. In an actual situation in which where *n* is small (as in nearly all experiments), the anonymity and no-communication conditions imposed on the experiment would be highly unusual.

Consider, for example, incremental value of a pristine rather than trash-littered beach against the slight inconvenience to any individual in carrying her own trash to a basket. Relative to this the social gain from cooperation in small group experiments is quite trivial. With *n* = 4, *a* = .3, perfect cooperation yields only a 20% improvement in payoffs relative to no cooperation at all.

So it is not a surprise (in terms of NSNX: for the standard theory, as mentioned earlier, it did come as a surprise) that propensity to cooperation increased in moving from IWW4.3 to IWW40.3 (from *n* = *4* to *n* = 40, with *a* = .3 in both cases). And in contrast to the robust tendency for contributions to decline over the 10 rounds, cooperation here did *not* regularly decline. For about half of the twelve IWW *n* = 40 groups, contributions did not significantly decline at all across rounds, and the decline even in the least successful groups was never close to complete free-riding. Baseline trials IWW ran for real money had already shown that behavior in *n* = 4 games did not noticeably depend on this. But it is interesting to note that the second most cooperative *n* = 40 group was a group playing for real money. And the most cooperative group of all was a group playing for real money with experience (having already played the game once). This is consistent with NSNX and seems sharply at odds with the many attempts to account for contributions as errors.

Another test also yields a striking result. The *lastround* choice is a player's last chance to move toward NSNX "neither selfish nor exploited" equilibrium. Figure 4 displays *lastround* choices for the n = 40 groups conditional on players' relative prior contributions over prior rounds.

Set DIFF = (*last round contribution) – (average contribution for the prior n-1 rounds)*

Gavg = mean own contribution up to the last round, relative to others' contributions*

So if a player gives more in the final round than her average prior gives, DIFF is positive, and if she gives less DIFF is negative. And *Gavg** is the grand average for *gavg* over the first n-1 rounds. If *Gavg** > 1, a player has given more on average than others, and if *Gavg** < 1 she has given less. Figure 4 then plots DIFF on the vertical axis against *Gavg** on the horizontal.

FIGURE 5 HERE

You can see a dichotomous response, in which about half the players who gave more than others gave nothing (the streak slanting down from the left), while the other half mostly *increase* their contribution over their prior round average. The figure shows a blurred "zero" line slanting down since player average gives across the 12 groups are not identical, so while the DIFF of a 0 final give must grow as *gavg** grows it is only roughly not exactly in proportion across players.

The simplest interpretation of the dichotomous result is that it is akin to the response we all have to gestalt drawings, like the duck/rabbit, young girl/old hag, faces/vases. In each case there are two ways to see the drawing. But at any moment we see just one. For players in the IWW small-n trials, this is not very significant. A person who has contributed conspicuously more than others will not have done much better from participating in the game than if she had been allowed to just opt out of the game. In fact she may easily have done worse, ending up with a net payoff less than her endowments. So we should expect a definite tendency to give less in the final round than whatever her average was prior to the final round, and a great majority of players do so.

But in the large-n, a = .3 tests, even players who contribute the most do very well. Consequently, in contrast to the n = 4 case, here two gestalts are available. The "exploited" gestalt which alone is easily prompted in the n = 4 case is available here, but also a "success" gestalt enhanced by reciprocity towards other contributors. Both make sense, as the gestalt drawing can reasonably be seen as either duck or rabbit. Heavy contributors have indeed been exploited by low contributors. But they have done very well anyway. And they cannot punish their exploiters without also punishing their fellow contributors. So it is not surprising (in hindsight: I do not claim to have foreseen this) that we would see the dichotomous response the scatter-plots reveal.

Putting this duck/rabbit interpretation in terms of the NSNX equilibrium (W = G'/S'): If weight to self-interest (W) is salient, from Rule 2 it must look large relative to a player's starting expectation. Player has given clearly more than others, and giving nothing in the final round will be the salient choice. But if it is your gain that is salient, then you see a big social gain (you and others who have been generous now have many more tokens than you were originally given). So G'/S' is large, and giving generously in the final round remains salient. And since some see the W 'duck' while others see the G'/S' 'rabbit, we can get the dichotomous response which is so apparent in Figure 4.

But it is a cognitive point, not a rational choice point of any sort, that we are prone to such either/or, duck/rabbit dichotomies: here with choice dominated by one side or the other of the equilibrium, not a balance between the two when (in this case) the shifts push in competing directions.

The next point, which is certainly important can't be seen in Figure 4 and has been essentially ignored in the very extensive literature on Public Goods games. Contrary to what might be expected on the standard theory, *complete* free-riding is far less common in the large group, *a* = .3 setting than in the small groups with the same return. Of 68 players in IWW4.3, seven were complete free-riders, never contributing at all. But of 480 players (from the same subject pool) in IWW40.3, there were only six complete free-riders.

This is the first of several occasions where we will encounter quite sharp reason to doubt that "types" (conditional cooperators vs. self-interested vs. full cooperators) often reported in analyses of experimental data are permanent types rather than transient and tactical. We have an almost 10-fold difference in propensity to be a complete free-rider which here cannot possibly be anything but an effect of the change in conditions.

But a puzzle can be seen in the large group response to an improvement in the per-token payoff. Players in the n = 4 trials were sensitive to the return they got per token in the pool. Raising return from .3 to .75, which lowers the net cost of contributing a token from .7 to .25, made a big difference. This remained true for n = 40 but only with respect to a difference in cooperation between a = .03 and a = .3. This tenfold increase in return -- reducing the net cost of giving a token from .97 to .7 -- produced a more than 10-fold increase in cooperation. This is not hard to understand, since with the very low return of .03 as few as 7 (out of 40) free-riders would leave most contributors as exploited losers. But a further increase from .3 to .75 in the n = 40 case (and also in the several n =

100) trials produced no increase at all in contributions. Players contributed no more when it cost only .25 tokens to give a token (a = .75) than when it cost .7 (a = .3).

The results grossly violate the one generalization in economic theory routinely designated a “law”: that lower price yields higher demand. The Law of Demand is suspended here, and absent any special conditions which make sense of such an observation in a few other contexts. But this insensitivity to quantitative effects relative to qualitative (for the large number case, even exploited players are doing very well whether n = 40 or n = 100) is consistent with the insensitivity to quantitative effects relative to qualitative already seen repeatedly. It is more evidence of insensitivity of choice to quantitative cues in unfamiliar contexts, which looks like an aspect of the strength of neglect defaulting central to the account of the cognitive illusions of Chapter 6. And it is yet more indication that any rational choice account of social choice, including NSNX, is likely to get into trouble unless serious allowance is made for cognitive effects.

Standard vs. Best Shot vs. Weakest link games

Now consider to a more intricate dataset from Croson, Fatas & Neugebauer, 2006. The CFN starting point was a standard 10-round Public Goods game, run with a surprise (to the players) restart after round 10. In all, the CFN series provides data on the basic game and two variants of the basic game, each played under three payoff conditions, and with new groups of players in each of the nine variants. Figure 5 shows the round-by-round results for the nine games, each played out for 20 rounds with a surprise restart at round 11.

FIG 6 HERE... 3 figures which shld be printed to facilitate comparisons, perhaps landscape. In the file they are just (sloppily) vertical, since I cdn't get them to line up neatly horizontal, which is probably what we want.

The baseline VCM game (first panel of figure 5) uses the “voluntary contribution” mechanism of a standard Public Goods game. The two variants use a Best Shot mechanism (BSM) where payoffs depend solely on the *highest* contribution in a group, or a Weakest Link mechanism (WLM) where payoffs depend solely on the *lowest* contribution (Hirshleifer & Harrison 198?). There are always four members in a group, an endowment of 50 tokens per round, and a return arranged to be the same across all variants when everyone makes the same choice. In VCM a player gets 2 tokens per token in the pool for the *average* contribution,/8 in BSM for the *highest* (Best Shot) contribution, and in WLM for the *lowest* (Weakest Link) contribution. A player's payoff for the round then is then the sum of what he gets from the pool plus any tokens kept (not contributed to the pool). If all players give the same, they all get the same payoff under any of these schemes, since then the average, highest and lowest *give* are all the same.

The three payoff conditions were: (1) the games just described, with no adjustment of payoffs; (2) an exclusion game (VCM-EX, BSM-EX, WLM-EX) in which the player (and ties unless all *gives* were identical) who gave the least forfeits any payoff from the pool; and (3) an exclusion with redistribution game (VCM-EX-R, BSM-EX-R, WLM-EX-R) in which any forfeited payoff is divided among the remaining players. In figure 4 the VCM, VCM-EX, and VCM-EX-R results are in the panel at the top, BSM, BSM-EX, and BSM-EX-R in the middle; and WLM, WLM-EX, and WLM-EX-R at the bottom.

I had occasion to mention at the start of this chapter that the bounce-back at the restart in the VCM game is a robust feature of Public Goods experiments. It always happens. This is a mystery for accounts in terms of self-interest, where confusion must play a large role in accounting for contributions. For it is certainly puzzling that a verbal instruction to start over and play another 10 rounds would return players to whatever state of confusion they were in before play started at all. The replays always look very similar to the original sequence, as they do here. One experiment (Cookson 2000) even followed a surprise restart with another surprise restart and then yet another surprise restart. Each restart produced a pattern similar to what is seen for the basic game here, with the repeated restart sequences looking much like the original sequence.

So we want to give an account of that, and more generally we want to use this interesting array of variants to consider how we might gain insight into what is happening in a poorly understood situation from what we can learn from how choices went in a related but better understood, situation. Here I want to move from the most familiar game (VCM) to increasingly harder-to-analyze games, taking advantage of what we have learned from the results of prior games. The VCM game is by now familiar. The Best Shot game is unfamiliar, but relative to the Weakest Link game is easy to analyze. We save the hardest for last.

Start from some general points about how the elements in the NSNX equilibrium (W, G', S') would respond to the changing incentives across the CFN variations.

From NSNX Rule 2, which conditions W (weight to self-interest) on a player's contribution relative to others in the group, W will be smaller to the extent that others' past and perhaps anticipated future cooperation become larger.

G' will reflect both a direct effect of a choice on payoffs to the group within the current round, and also (until the last round) possibly some adjustment for effect of a choice on the prospects for enhanced payoffs in future rounds.

S' will have a direct cost offset by any expected return from the pool, and perhaps also by a share in any prospective value for G'. When the CFN exclusion conditions are in play, the effective cost of contributing (*net* S') needs to be also adjusted for the prospect of exclusion. The higher the *give*, the less chance of exclusion. Giving 50 guarantees against exclusion, since 50 cannot be less than others have contributed. Giving 0 guarantees exclusion. No one could give less./8

Against that background. consider what NSNX implies with respect to the results you see in Fig. 5.

VCM vs. VCM-EX vs. VCM-EX-R. The VCM results are in the lowest line of the first panel. They are entirely "normal" for a Public Goods game with a return/token as high as .5. The conspicuous restart effect is also entirely normal, as already stressed, and has a ready NSNX interpretation. Players are left frustrated by the modest level of cooperation by the 10th round and readily take up a suggestion to try again. And on that explanation of the restart effect, of course there is no reason for a restart effect when cooperation is succeeding not failing, and the Best Shot and Weakest Link data in the other panels reflect that.

Between VCM and VCM-EX, on any reasonable account (and with no need to allow for other-regarding motivation) contributions should move quickly toward 100%. If they didn't (but you can see in figure 5 that they do) an account explaining some severe cognitive distortion would be needed. In VCM a player who gives 0 gets a better payoff than a player who gives any positive amount. The more you give, the less you get, providing the tension between self-interested and social motivation that is the heart of the game. Unless others are also giving, a player motivated by "neither selfish nor exploited" concerns does not want to give. A player motivated only by self-interest does not want to give at all. But in VCM-EX, a player who gives 0 guarantees he will get nothing from what promises to be a very profitable pool (no one could give less). A player who gives a positive amount but not his full endowment (50) risks ending the round with a loss (relative to his endowment). He has moved tokens from his private account to the pool, but might be excluded and get nothing back from the pool.

But a player who gives 50 loses only if the other three players together do not at least match that, which in fact rarely happens in round 1 of any Public Goods game, and becomes especially unlikely to happen under the exclusion rule. And even in that never-occurring worst case, the player has risked only 25 tokens (since half his 50-token contribution will come back) to enhance the prospect of gaining 10 x 50 tokens over the run of the game. It is not surprising that at least one player in every group in these trials gives 50. But then a player who gave < 50, but wants to do better than just keep his endowment, must enter a kind of reverse auction against other players who want to share in the pool, where the only safe choice is to give 50. Self-interest alone is then enough to drive contributions up towards 100%. So it is scarcely surprising that contributions are high from the start, indeed push towards 100% compliance, and with no restart effect since in this game cooperation does not fail.

And under VCM-EX-R (exclusion + redistribution) the prospect for complete cooperation becomes overwhelming. Only in some very improbable condition could it fail to be profitable, while giving zero means getting nothing from the pool and giving a positive amount less than 50 risks exclusion. Even if no one else contributed, giving the entire endowment is profitable, since if others give 0 the sole contributor collects all earnings from the pool, which doubles her contribution. But given the strength of the VCM-EX > VCM inference there could hardly be room for much more improvement. Overall we should expect VCM-EX-R > VCM-EX >> VCM. Further, the bizarre choices seen in Chapter 7 for the inverted Public Goods game, defying the logic of the game, can hardly occur here though there is no reason to suppose these players would do any better in the Chapter 7 games. For the threat of exclusion (or experience of exclusion, since at least one player will be excluded until all are fully contributing) pushes choices in one direction – up – a push lacking in the Chapter 7 situations.

Finally, by way of an exercise in the mechanics of the NSNX formalism, notice that *net* S' for *give* = 50 becomes negative under VCM-EX as expected contributions rise (giving 50 becomes essentially certain to increase own-payoff). G'/S' becomes negative, hence necessarily W > G'/S', since W is positive by definition. So the see-saw (fig. 2 in Ch. 1) must tilt towards self-interested choice, though all three elements of the NSNX equilibrium have changed to favor group-interest in VCM-EX relative to VCM. But if S' is negative (it pays to give rather than costs to give) and G' positive, then there is no conflict between group- and self-interested choice. *Give* = 50 is best for the group, but it is also best for self-interest, so the effect through the NSNX equilibrium is odd but not perverse. This technical point has already been noticed in the discussion of Chapter 1.

BSM vs. BSM-EX vs. BSM-EX-R. If payoff from the pool scales with either the average (VCM) or smallest (WLM) contribution, then the socially best outcome is for everyone to contribute their entire endowment, yielding each player double their endowment. But under the *Best Shot* rule only one player need contribute a full endowment to give everyone the maximum payoff from the pool (double the endowment). Any contribution beyond one person contributing the maximum is pointless. If players could communicate, they would arrange for one full contribution and three zero contributions, ending up with payoffs almost triple their endowment./9 But no communication is allowed, so players must gamble.

Once exclusion is added to the best-shot game the analysis is far simpler. With exclusion, giving 0 guarantees getting nothing from a pool that (if 50 were given) guarantees a return of 100. Choosing any number in the interval between 0 and 50 risks exclusion. That forces the inverse auction noticed in VCM-EX, to avoid being the low contributor, but in even starker form since no matter what others do giving 50 guarantees getting 100 from the pool, and giving anything less than 50 risks getting nothing from the pool. It is not surprising on any account (NSNX, pure self-interest, or whatever) that we see almost 100% contributions from the start and literally 100% contributions in almost all remaining rounds. The clear comparative statics inference is that BSM-EX >> BSM, as indeed is seen in figure 4.

But BSM-EX-R (best shot with exclusion + redistribution) yields a surprising result. In terms of self-interest, the incentive to contribute 50 is even stronger than in BSM-EX (exclusion *without* redistribution). For there is a chance someone will give < 50, be excluded, and leave his share to be divided among those not excluded. The incentive to contribute 50 is a bit stronger. But the result is the reverse. The top line in the Best Shot panel is for BSM-EX. In defiance of players' self-interest, the BSM-EX-R result never exceeds the BSM-EX result, but instead falls below it in almost every round./10

This makes no sense for self-interested players, who would see only one effect of redistribution. Expected payoff (if she gives 50) is slightly higher. Consequently, in terms of self-interest the results seen in Fig. 5 should not happen. If, however, NSNX motivation is in play then the result is correct. For with NSNX motivation there are two effects, not one from giving 50. There is the self-interested effect already noticed, strengthening the incentive to give 50. But there is also a group-interest effect, since exclusion no longer lowers overall group payoff. In BSM-EX, a low contributor's payoff from the pool disappears. In BSM-EX-R that payoff is not lost but only redistributed.

Consequently, adding redistribution yields a pure NSNX result. Under BSM-EX, give = 50 is the only sensible choice for self-interest, and group-interest reinforces that. So BSM-EX must yield more cooperation than BSM alone (no exclusion) even without NSNX, but NSNX reinforces that. But in BSM-EX-R the group-interested reinforcement of self-interest favoring *give* = 50 no longer holds. With redistribution, exclusion can never entail any loss whatever to group payoff. Rather total group payoff is improved. In the Best Shot game, any contribution beyond one 50-token gi*ve* can yield no increase in payoff from the pool. If a player is excluded by his low *give*, but his payoff is not lost but only redistributed (what happens in BSM-EX-R), total group payoff *increases* by avoiding a redundant contribution.

A perfectly group-interested player would be willing to make that sacrifice. But it is hardly conceivable that a NSNX player (who attends to self-interest as well as group-interest) would. But even a strictly self-interested player confused or uncertain about how the game works, or revealing a lack of concentration, or caught by the here inappropriate free-rider default in the cascade, might give less than 50 and be excluded.

But that chance would not increase in BSM-EX-R relative to BSM-EX. But with NSNX motivation, losing the group-interest component of the incentive does making give < 50 more likely. That is the clear NSNX inference, though without the results we can see in figure 4, we could hardly have expected an effect this subtle to be strong enough to be noticed. But in figure 4 that effect, though small, is nevertheless large enough to be very easily seen. That is probably because another player – not a player making a mere blunder -- could recognize (or in an earlier round notice) that someone might give 0, at least one other give 50 (essentially certain here), so creating a case such that a minimal 0 < *give* < 50 contribution might both share in the group payoff and save some endowment (save 49 of 50 tokens if *give* = 1). And this would improve *both* own payoff and aggregate group payoff.

Further, even if this looks like (and indeed is) a rather reckless gamble, it is significant for NSNX that it is a risk only for the chooser himself. If he feels like taking a fling it is only his own payoff he is putting at risk. From the group perspective, it is a nice thing to do.

So NSNX suggests more deviation from give = 50 in BSM-EX-R compared to BSM-EX, while self-interested motivation alone does not. The incentive to be alert is softened. And even if fully alert, a player who feels like fooling around in some rounds would see no social harm in it. Rather, it is socially efficient. Hence in contrast to VCM and (as will be seen next) WLM, adding redistribution to exclusion does not imply higher propensity to contribute, but marginally the opposite. This is a surprising turn in the logic. But in the data in figure 4 the effect turns out to be strong enough to be readily apparent.

WLM vs. WLM-EX vs. WLM-EX-R. WLM is the same game, differently presented, as the Minimum game discussed in Chapter 7./11 Cooperation is modestly more successful here. Groups do not quickly sink to their lowest payoffs. Contributing 0 in the Weakest Link game here, as the note explains, is the same as coordination on '1' in the Minimum game. The difference in results (poor in WLM rather than downright terrible in the Minimum game) is most easily accounted for by the small size of the CFN groups (4 players) relative to Minimum game experiments (usually at least 7 players). As

I mentioned in Chapter 7, I will eventually try to show that there is a significant *cognitive* puzzle in why players do so badly in these games. But I continue to set that aside. We will be only concerned here with comparative statics across the WLM vs. WLM-EX vs. WLM-EX-R games. Given the results of the WLM game, what should we expect in WLM-EX and WLM-EX-R? And if a puzzle arises, can NSNX resolve it?

For any player, define *MIN* as the lowest another player chooses. Under WLM, a choice < *MIN* lowers own-payoff and everyone else's as well, a choice > *MIN* wastes tokens, yielding nothing to either the chooser or anyone else. So it is not a puzzle that over the sequence of rounds contributions tend towards settling on some common level, with the minimum rule biasing that toward the low end of the range of initial responses.

With exclusion, incentives change but whether cooperation should increase is unclear. WLM-EX incentives are dichotomous. At equilibrium no one gives anything is possible (and everyone just keeps their endowment). Everyone gives the maximum is also possible (no one is excluded and everyone gets the maximum payoff). But there is no viable option in the middle. A player who moves up puts pressure on lower players to either move up (otherwise they will be excluded) or move down to 0 (which keeps a chooser's endowment intact, but gives up on any prospect of making a profit from the game). A player who moves down (and he should move down to zero, since being *at* the minimum otherwise makes no sense at all) pressures everyone else to follow, to everyone's disadvantage including the initiator.

With exclusion, then, the minimum within a group must be expected to move *either* move up towards full cooperation or collapse to no cooperation. The intermediate levels in for WLM-EX (and EX-R) are averages across all players. Within any group, though, the result is that either it moves up towards full cooperation or down toward no cooperation. In terms of self-interest alone, risk-aversion would seem to make moving down salient. It takes only one choice = 0 to make a positive give costly to anyone else and everyone knows that. The experimenters expected that. But exclusion of the low giver would also provide a player who started from a middling level a clear incentive to give more in future rounds. And if motivation is NSNX, punishing the least cooperative player is socially benign in a context where cooperation could double payoffs for everyone, as indeed even a player risking punishment is likely to allow. So although it is unclear how these effects would net out, in terms of NSNX it is certainly not surprising that cooperation can in fact be observed to increase. That can be seen in the first series (rounds 1 – 10), and it becomes clearer as players with experience from rounds 1-10 are given a chance to try again in rounds 11-20.

In turn this yields a stronger inference for WLM-EX-R. Although NSNX yields no clear prediction about WLM-EX relative to WLM, it does imply WLM-EX-R > WLM-EX. So seeing that adding exclusion improves cooperation, adding exclusion + redistribution we can expect WSM-EX-R >> WSM. For suppose that in WLM-EX-R a *give* leads to exclusion. That is no worse for self-interest than in WLM-EX. But if it doesn't, a player probably does better and never does worse than in WLM-EX. And group-interest is clearly better-served in WLM-EX-R relative to WLM-EX, since everyone other than the excluded gets a bonus and a person excluded is no worse off than under WLM-EX. A strong point, however, is that NSNX motivation softens the severity of the choice. A middling choice might result in exclusion, but for a player with "neither selfish nor exploited" motivation, it is not so selfish as giving 0, and not so vulnerable to feeling exploited as giving 50 (if the prior round minimum was far below that). And the exclusion is socially fair and socially useful (recalling the discussion of contributors to G' in Chapter 1). All this must make a player tempted to bail out (choose 0) more willing to take a chance on a middling level, and a player at a middling level in the prior round more likely to head up rather than head down.

So the NSNX inference is that players should do better under WLM-EX-R than under WLM-EX, which is better than WLM. And the data in fig. 4 show that players do better, and very strikingly they do better when given an opportunity to replay the game (rounds 11-20) with experience.

So reviewing the entire CFN set, the basic Public Goods games show one familiar but never actually explained effect (the bounce-back at restart) which makes easy sense in terms of NSNX motivation. Adding *exclusion* and *exclusion + redistribution* to Voluntary Contribution yielded almost perfect cooperation, but in a way that is easily explained even in terms of self-interest. NSNX motivation only reinforces what self-interest alone can explain. But adding *exclusion* and then *exclusion + redistribution* to the Best Shot and Weakest Link games provides results (subtle for the Best Shot games, impossible-to-miss for the Weakest Link games) which no available model other than NSNX seems to make sense of.

But other than the dichotomous result of the lastround test (figure 3) cognitive effects have not played much of a role in this chapter. The surprising results in the Best Shot and Weakest Link games get an explanation which is essentially pure NSNX. That will change in the next two chapters, where interactions of cognitive effects with NSNX are prominent throughout.