9-1-2012

Identification in games : changing places

Darrell Patrick ROWBOTTOM
Lingnan University, Hong Kong

Recommended Citation

Rowbottom, D. P. (2012). Identification in games: Changing places. Erkenntnis, 77(2), 197-206. doi: 10.1007/s10670-011-9357-0

Preprint – Forthcoming in *Erkenntnis*

Identification in Games: Changing Places

Darrell P. Rowbottom

Department of Philosophy

Lingnan University

DarrellRowbottom@ln.edu.hk

Abstract

This paper offers a novel 'changing places' account of identification in games, where the consequences of role swapping are crucial. First, it illustrates how such an account is consistent with the view, in classical game theory, that only outcomes (and not pathways) are significant. Second, it argues that this account is superior to the 'pooled resources' alternative when it comes to dealing with some situations in which many players identify. Third, it shows how such a 'changing places' account can be used in games where some of the players identify with one another, but others do not. Finally, it illustrates how the model can handle the notion that identification comes in degrees.

1. Identification and the Prisoner's Dilemma

Consider a simple Prisoner's Dilemma scenario, as represented below in fig. 1. Classical game theory suggests that either player ought to select *defect*, although empirical studies indicate that many people—in fact, only marginally under 50% on the average over the studies considered by Sally (1995)—select *cooperate*.[1] So rather than accept the implausible consequence that a significant percentage of people are irrational, authors such as Bacharach (2006) invite us to conclude instead that some form of identification is occurring between the players.

Fig. 1—Simple Prisoner's Dilemma

		Player Two	
		Cooperate	*Defect*
Player One	*Cooperate*	2, 2	0, 3
	Defect	3, 0	1, 1

The purpose of this paper is not to explain why identification occurs—that is, why individuals prefer cooperative strategies in some circumstances—but to propose a new way to *understand, or model, the mechanism* by which it occurs. So in the subsequent discussion, we will take it as given that identification occurs for reasons independent of the games under consideration.

The first hurdle is to decide whether it matters how a game is played. Identification might, for instance, merely consist of empathy with respect to the experience of being cheated or defeated in some game; the disutility, that is to say, of being on the receiving end of ungenerous, or downright selfish, behaviour. Alternatively, identification might be concerned purely with the outcome. I take this to be preferable because the model is kept simple; we need not worry about pathways to consequences rather than consequences themselves. As we will see, however, there is more than one model of identification that respects this requirement.

[1] Fear of reprisal is ruled out in studies where neither player knows the identity of the other, and both know that they will only play the game with one another once.

Now my basic idea is that we can model identification, say of player one with player two in fig. 1, by imagining that player one understands that he might find himself in player two's shoes *with respect to the outcome*. So we can imagine him knowing that he will make the choice of *cooperate* or *defect* as player one, but that he will then have a half chance of swapping places with the other player (i.e. becoming player two) at the conclusion of the game.

Fig. 2—'Changing Places' Perspective of One Player Identifying with Another

		Player Two	
		Cooperate	*Defect*
Player One	*Cooperate*	2, 2	3, 0 or 0, 3 (P=0.5 each)
	Defect	3, 0 or 0, 3 (P=0.5 each)	1, 1

Intuitively, opting for *defect* is now the less attractive option (even if the other player sees the potential results differently).[2] But is the probability supposed to be understood as aleatory, i.e. to reflect a propensity or a statistical frequency? Or is it supposed to be understood as personal, e.g. as a coherent degree of belief? Taking inspiration from the idea of the veil of ignorance, in the version of Vickrey (1945) and Harsanyi (1953, 1955), I opt for the latter.[3] To identify with someone else, in the sense proposed here, is to behave *as if* one might end up in their shoes (and they might end up in one's own shoes). Being given the possibility of ending up in any position in society, *and assigning each possible outcome an equal*

[2] I will not, here, try to fully explain that intuition; this is a topic explored by Sugden (1993, 2000) and Bacharach (2001, 2006).

[3] Rawls (1971), who is perhaps most often associated with the 'veil of ignorance' idea, wanted to avoid the use of probabilities in decision making; instead, he proposed to use a maximin decision rule. This is explained and criticised by Harsanyi (1975).

probability, leads one to consider carefully how best society should be organised for its citizens. Similarly, I contend, considering a game from such a perspective leads one to consider how best the game should be conducted for its participants.

This raises the question of whether it is rationally required to assign each possible outcome an equal probability, e.g. to follow the principle of indifference advocated by Keynes (1921) or the modern equivalent, the maximum entropy principle, advocated by Jaynes (2003) and Williamson (2010). The basic idea behind such principles is that one should equivocate over possibilities in the absence of any information to indicate that any particular possibility, rather than any other, obtains (or will obtain); and the difficulties with this idea have been discussed extensively, e.g. by Seidenfeld (1986), Gillies (2000), and Rowbottom (2008, 2011).

However, it is *not* necessary for such a rule to hold in order for the 'changing places' account to be tenable. Instead, it may be proposed simply that (full) identification *involves* assigning equiprobability across possible place changes (and the possibility of not changing places). So in an n+1 player game where a given player fully identifies with the other n players, his 'changing places' probability (or swap probability) for each other player will be $1/n$.[4] And if that condition fails to obtain, we must conclude that the given player does not fully identify with the other n players. As we will later see, this opens the door for modelling *degrees* of identification too. (These seem to exist given that it is intuitively plausible to say that one might identify rather more with one player than another, yet still identify with both to some

[4] A game with finitely many players is assumed throughout; the model will break down for infinitely many players, because zero probabilities may nonetheless correspond to possibilities in such circumstances. See, for example, Williamson (2007).

extent, and so forth.) In essence, the idea is that swap probabilities may differ. But for the moment, until section 4, take all mentions of 'identification' to refer to full identification.

2. 'Changing Places' versus 'Pooled Resources' Approaches

Naturally, there are alternatives to the 'changing places' approach outlined above. One notable example is to take what might be called a 'pooled resources' perspective. Here, as depicted in fig. 3, the player identifying with the other would be said to imagine that the overall utility from the outcome will be shared equally.

Fig. 3—'Pooled Resources' Perspective of One Player Identifying with Another[5]

		Player Two	
		Cooperate	*Defect*
Player One	*Cooperate*	2, 2	1.5, 1.5
	Defect	1.5, 1.5	1, 1

Again, the result appears to be intuitively desirable. Clearly *cooperate* is the best option for the player with a perspective such as that outlined in figure 3. However, finding a justification for modelling typical situations of identification in such a fashion is not easy; there is no obvious analogue of the veil of ignorance in the 'changing places' model. Moreover, the pooled resources view leads to undesirable results, as we shall see below, when somewhat more complicated games are considered.

[5] Taking a 'team perspective', i.e. considering combined utility, results in a similar ranking. The payoff at the level of the team is 4 if both players cooperate, only 2 if both players defect, and 3 if one player defects and the other cooperates.

Consider now a three-way scenario akin to the classic Prisoner's Dilemma, as depicted below in figure 4. With no identification present, the best option for any player is clearly to *defect* rather than *cooperate*.

Fig. 4—Three Player Prisoner's Dilemma Scenario

		Player Three	
		Coop	*Def*
Players One, Two	*Coop, Coop*	2, 2, 2	0, 0, 3
	Coop, Def	0, 3, 0	0, 3, 3
	Def, Coop	3, 0, 0	3, 0, 3
	Def, Def	3, 3, 0	1, 1, 1

Think now, however, about how the situation looks from the perspective of a player who identifies with the other two on the 'pooled resources' model of identification:

Fig. 5—Three Player Prisoner's Dilemma Scenario for a 'Pooled Resources' Identifier

		Player Three	
		Coop	*Def*
Players One, Two	*Coop, Coop*	2, 2, 2	1, 1, 1
	Coop, Def	1, 1, 1	2, 2, 2
	Def, Coop	1, 1, 1	2, 2, 2
	Def, Def	2, 2, 2	1, 1, 1

Here, *cooperation* is no more appealing than *defection*. (Imagine player three is the 'pooled resources' identifier. *Cooperation* gives two chances for 2 and two chances for 1; *defection* is indistinguishable.) But on a 'changing places' perspective of identification as depicted below (fig. 6), by way of contrast, we still achieve the desirable result that *cooperation* is the more attractive option provided that the player is risk averse (i.e. she would rather have '2 in the hand' than a third of a chance of ending up with nothing and two thirds of a chance of ending up with 3):

Fig. 6—Three Player Prisoner's Dilemma Scenario for a 'Changing Places' Identifier[6]

		Player Three	
		Coop	*Def*
Players One, Two	*Coop, Coop*	2, 2, 2	0, 0, 3 or 0, 0, 3 or 3, 0, 0 (P=1/3 each)
	Coop, Def	0, 3, 0 or 3, 0, 0 or 0, 3, 0 (P=1/3 each)	0, 3, 3 or 3, 0, 3 or 3, 3, 0 (P=1/3 each)
	Def, Coop	3, 0, 0 or 0, 3, 0 or 0, 0, 3 (P=1/3 each)	3, 0, 3 or 0, 3, 3 or 3, 0, 3 (P=1/3 each)
	Def, Def	3, 3, 0 or 3, 3, 0 or 0, 3, 3 (P=1/3 each)	1, 1, 1

Admittedly, we would like not to have to assume that the identifying player is risk averse; however, this model still trumps the 'pooled resources' one because in that event all the results are 'in the hand'.

3. Identification of Sub-Groups on a 'Changing Places' Model

The fruits of the 'changing places' model of identification may also be seen by considering a scenario in which players two and three identify with one another, but neither identifies with player one. This is a version of the Three Player Prisoner's Dilemma where two of the prisoners are partners in crime, but the third stands alone. (Perhaps he is not part of the gang that the other two are in. Maybe he is a driver that was hired for the botched bank job that resulted in the arrest of the three players.)

[6] I should mention another possible model, where each possible change of place is taken into account, and multiple swaps are allowed. In the three player case, for example, one would consider six possible orderings: 1, 2, 3; 2, 1, 3; 3, 2, 1; 1, 3, 2; 2, 3, 1; and 3, 1, 2. Clearly, such a model is considerably more complex than the one considered here. It is conceivable, however, that it may be superior in some situations.

Fig. 7—Three Player Prisoner's Dilemma Scenario for a 'Changing Places' Identification between Players Two and Three (and/or *vice versa*)

Players One, Two		Player Three: *Coop*	Player Three: *Def*
	Coop, Coop	2, 2, 2	0, 0, 3 or 0, 3, 0 (P=1/2 each)
	Coop, Def	0, 3, 0 or 0, 0, 3 (P=1/2 each)	0, 3, 3
	Def, Coop	3, 0, 0	3, 0, 3 or 3, 3, 0 (P=1/2 each)
	Def, Def	3, 3, 0 or 3, 0, 3 (P=1/2 each)	1, 1, 1

As indicated, the best option for two and three is to both defect; that way, they are both protected from losing out altogether (and also have the prospect of both gaining as much as the game allows in principle, in the event that player one decides, foolishly, to cooperate). This is the result that we want.[7]

It is true that a similar result *can* be gained from the 'pooled resources' model of identification—see fig. 8, below—but we can already seen why this model is inferior in other circumstances.

Fig. 8—Three Player Prisoner's Dilemma Scenario for a 'Pooled Resources' Identification between Players Two and Three (and/or *vice versa*)

Players One, Two		Player Three: *Coop*	Player Three: *Def*
	Coop, Coop	2, 2, 2	0, 1.5, 1.5
	Coop, Def	0, 1.5, 1.5	0, 3, 3
	Def, Coop	3, 0, 0	3, 1.5, 1.5

[7] Identification need not be taken to be reciprocal in the account offered here. We can also imagine a case where three identifies with two but two does not identify with three; then three considers the situation as depicted in fig. 7, but one and two consider the situation as depicted in fig. 4 (provided neither identifies with the other).

	Def, Def	3, 1.5, 1.5	1, 1, 1

4. Extension of the 'Changing Places' Model to Cover Degrees of Identification

A further advantage of the 'changing places' model, which was briefly mentioned in the introduction, is that it provides a natural means by which to cope with the notion that identification comes in degrees. Consider first a simple two player Prisoner's Dilemma from player one's perspective.

Fig. 9—Possible Degrees of Identification of Player One with Player Two in a Prisoner's Dilemma

		Player Two	
		Cooperate	*Defect*
Player One	*Cooperate*	2, 2	3, 0 (P=n) or 0, 3 (P=1-n)
	Defect	3, 0 (P=1-n) or 0, 3 (P=n)	1, 1

If n is one half, then full identification occurs. If n is zero, then no identification occurs. And if n is greater than zero but less than one half, then partial identification occurs.[8] (What if n is higher than one half? We will come to this shortly.) Thus there is a specific range of degrees of identification for which *defect* is a preferable option to *cooperate* (and *vice versa*). One may consider an expected utility calculation in order to see this. Imagine, for instance, that player one thinks that player two is equally as likely to *cooperate* as to *defect*; P(two *cooperates*)=P(two *defects*)=0.5. The expected utility of selecting *cooperate* is 0.5*2 + 0.5*3n, whereas the expected utility of selecting *defect* is 0.5*3(1-n) + 0.5*1, from player

[8] One need not think that there are precise numerical degrees of identification (just as one need not think that there are precise numerical degrees of belief); in principle, one may work with imprecise probabilities instead.

one's perspective. Setting these equal, we can find the point at which both selections have the same expected utility:

$$1 + 3n/2 = 3/2 - 3n/2 + 1/2$$

$$3n = 1$$

Hence when n is greater than a third and less than one half, *cooperate* is the preferable option for player one; when n is lower than a third, *defect* is preferable.

It is also possible to allow n to be larger than one half but less than or equal to one, in which case we might say that player one *overidentifies* with player two. In the extreme case where n is one, for example, player one is concerned *only* with how the outcome of the game will affect player two. He will therefore see *cooperate* as the preferable choice. (One might think it more accurate to say that full identification occurs when n is one, and that player one *equally identifies* with himself and player two when n is a half; that is, if one allows for self-identification. However, I prefer to adopt the current convention for 'identification' in order to avoid confusion.) Overidentification need not be irrational, although it may often be altruistic and supererogatory. [9]

The advantages of allowing for degrees of identification become even clearer when we consider games involving more than two players. To see this, let us again consider a Three Player Prisoner's Dilemma.

[9] Modifications to the 'pooled resources' approach might also be made in order to allow for degrees of identification; for example, a given player might assign weights to the utilities of the other players. The use of probabilities in the 'changing places' account makes the development natural, however, in a way that it has not previously been.

Fig. 10—Player One's Degrees of Identification in a Three Player Prisoner's Dilemma

		Player Three: *Coop*	Player Three: *Def*
Players One, Two	*Coop, Coop*	2, 2, 2	0, 0, 3 (P=n) or 0, 0, 3 (P=m) or 3, 0, 0 (P=1-m-n)
	Coop, Def	0, 3, 0 (P=n) or 3, 0, 0 (P=m) or 0, 3, 0 (P=1-m-n)	0, 3, 3 (P=n) or 3, 0, 3 (P=m) or 3, 3, 0 (P=1-m-n)
	Def, Coop	3, 0, 0 (P=n) or 0, 3, 0 (P=m) or 0, 0, 3 (P=1-m-n)	3, 0, 3 (P=n) or 0, 3, 3 (P=m) or 3, 0, 3 (P=1-m-n)
	Def, Def	3, 3, 0 (P=n) or 3, 3, 0 (P=m) or 0, 3, 3 (P=1-m-n)	1, 1, 1

If n has the same value as m, then player one fully identifies with player two; and if 1-m-n also takes the same value as n, then player one fully identifies with player three too. But player one may fully identify with player two while identifying with player three to a lesser extent, for example, in the event that $m = n > 1\text{-}m\text{-}n$. Then if the difference in degree of identification is great enough, he will prefer a different option than he would if he fully identified with both players. (Naturally, other orderings of identification are possible: $n > m > 1\text{-}m\text{-}n$; $n > 1\text{-}m\text{-}n > m$; $n = 1\text{-}m\text{-}n > m$; and so forth.) In general, it is possible to perform expected utility calculations, such as the one illustrated above, in order to determine which option player one should prefer for any given values of n and m.

Before concluding, there are two related questions that I should like to address. First, are there any reasons independent of the prior considerations to think that identification genuinely involves considering changing places? (And does it matter if it really does?)

Second, is the 'changing places' approach superior to the 'pooled resources' approach in general, i.e. in all possible applications?

5. The Reality and Scope of the 'Changing Places' Account

There may be some reason, independent of the prior considerations, to think that people genuinely 'put themselves in the shoes of others' when they're considering how they should behave. For example, one might appeal to the evidence concerning the existence of mirror neurons in humans—see Iacoboni et al. (1999)—and the simulation theory of mind championed by Goldman (2006). And although it is not plausible that people make rather complicated expected utility calculations in order to decide what to do, the *normative* aspect of game theoretical approaches must be borne in mind. In short, a theory of what one *ought to do* when one identifies may be correct in spite of what people actually do.

As an anti-realist, however, I think that whether the 'changing places' model is *realistic* is of little concern; I am interested in whether it saves the phenomena, i.e. gives good predictions, in an economical way.[10] So I commend it, on *pragmatic* grounds, as more theoretically virtuous than the available alternative (in, as I will explain below, a significant class of scenarios). In particular—to borrow from Kuhn's (1977) list of theoretical virtues—I have argued that it has more accuracy and scope than the 'pooled resources' approach, although it is somewhat less simple. (There is often a trade-off to be made between scope and simplicity; by introducing further complexity into one's models, one can handle more scenarios. There

[10] See Rowbottom (In Press).

are many examples in elementary physics, e.g. in the move from the ideal gas law to the van der Waal's gas law.[11]) So in summary, I would say that the 'changing places' model is a useful way of *thinking about* identification; it is a convenient fiction.

This brings me to the question of whether I have shown that the 'changing places' approach is *generally* superior to its 'pooled resources' rival. I have not, of course, because I have not had the luxury of considering all possible identification-based scenarios. Thus I concede that there may be some circumstances in which it is easier to use the latter than the former. (An analogy, given what I have said above, may help. No-one uses special relativity, rather than Newtonian mechanics, when doing biomechanics; the former is easier to use in such scenarios, although it is less accurate and has less scope.) I also concede that there may be some circumstances in which it is possible to use the latter but not the former—when unusual internal groups structures are present, for example—although I cannot presently conceive of any.

6. Conclusion

We have considered two different ways of modelling identification, via 'changing places' or 'pooled resources' approaches, and seen that the former is superior on a number of counts. (It should also be remembered that 'team perspectives' are similar to 'pooled resources' approaches; see footnote four.) It may be employed to favour cooperation (when risk aversion is present) in three (or more) player Prisoner's Dilemma scenarios with full identification,

[11] In the latter, but not in the former, the volume of molecules and the forces between molecules are taken into account.

and also be used to explain the preferences of identifying sub-groups in such games. Furthermore, crucially, it may also handle the notion that identification comes in degrees, and therefore that any given player in a game may identify with several other players to some extent, but with some more than others. And all this is achieved without abandoning the assumption that identification may be understood in terms only of the outcomes of (and not the pathways that occur in) games.

Acknowledgements

I am grateful to three anonymous referees for their comments on a previous version of this paper, which helped my thinking on these matters considerably. I should also like to thank the British Academy and the Sydney Centre for the Foundations of Science for supporting this research.

References

Bacharach, M. 2001. 'Superagency: Beyond an Individualistic Theory of Games'. In J. van Benthem (ed.), *Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge*. San Francisco: Morgan Kaufmann, 333–337.

Bacharach, M. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. N. Gold and R. Sugden, eds. Princeton: Princeton University Press.

Gillies, D. 2000. *Philosophical Theories of Probability*. London: Routledge.

Goldman, A. I. 2006. *Simulating Minds.* Oxford: Oxford University Press.

Harsanyi, J. C. 1953. 'Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking', *The Journal of Political Economy* 61, 434–435.

Harsanyi, J. C. 1955. 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility', *The Journal of Political Economy* 63, 309–321.

Harsanyi, J. C. 1975. 'Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory', *The American Political Science Review* 69, 594–606.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., Rizzolatti, G. 1999. 'Cortical Mechanisms of Human Imitation', *Science* 286, 2526–2528.

Jaynes, E. T. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Keynes, J. M. 1921. *A Treatise on Probability*. London: Macmillan.

Kuhn, T.S. 1977. 'Objectivity, Value Judgment and Theory Choice', in *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Belknap.

Rowbottom, D. P. 2008. 'On the Proximity of the Logical and "Objective Bayesian" Interpretations of Probability', *Erkenntnis* 69(3), 335–349.

Rowbottom, D. P. 2011. *Popper's Critical Rationalism: A Philosophical Investigation*. London: Routledge.

Rowbottom, D. P. In Press. 'The Instrumentalist's New Clothes', *Philosophy of Science*.

Sally, D. 1995. 'Conversation and Cooperation in Social Dilemmas: A Meta-analysis of Experiments from 1958 to 1992'. *Rationality and Society* 7, 58–92.

Seidenfeld, T. 1986. 'Entropy and Uncertainty', *Philosophy of Science* 53, 467–491.

Sugden, R. 1993. 'Thinking as a Team: Towards an Explanation of Nonselfish Behavior'. *Social Philosophy and Policy* 10, 69–89.

Sugden, R. 2000. 'Team Preferences', *Economics and Philosophy* 16, 175–204.

Vickrey, W. 1945. 'Measuring Marginal Utility by Reactions to Risk', *Econometrica* 13, 319–333.

Williamson, J. 2010. *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.

Williamson, T. 2007. 'How Probable is an Infinite Sequence of Heads?', *Analysis* 67, 173–180.