76. Informal agreements

Martin Dufwenberg

While economists devoted much effort to developing theories regarding the formation and impact of binding contracts, they gave far less attention to informal agreements. The main exception is when game-theoretic analysis has been complemented with an interpretation that certain equilibria correspond to (pre-play articulated) informal agreements struck by the players. See van Damme (1987, p.4), Catonini (2021), McCutcheon (1997), Baker et al. (2002), Levin (2003), and Jang et al. (2018) concerning equilibria (or related notions) in general, cartel agreements, relational contracts, and MOUs. Such an interpretation may be realistic if it refers to a setting where people talk, and it may be helpful for justifying why players might coordinate on an equilibrium in the first place. However, no presumption is made that the informal agreement, per se, changes players' motivation by having an effect on their utilities. Behavioural economists will recognise that this can be restrictive. Several experimental studies have shown that if subjects are offered an opportunity to enter an informal agreement, then they often do so and they also typically act in accordance with the agreement subsequently. See Malhotra & Murnighan (2002), Irlenbusch (2004), Ben-Ner & Putterman (2009), Kessler & Leider (2012), Krupka et al. (2017), Dufwenberg et al. (2017) (DSV). There is little related theory. Miettinen (2013) assumes that players have a preference for not reneging and shows that this can impact strategic play in some games. Then there is DSV, who propose a general approach for modelling both the formation and the impact of informal agreements.

A key building block in DSV's analysis is the notion of 'an underlying game', an extensive game form G which is taken to describe everything relevant to a strategic situation, except that the players are also assumed to get together and discuss and potentially strike informal agreements. DSV assume that the objects of the negotiations are strategy profiles in G. Predictions are then delivered via a solution concept with a special structure. Namely, DSV propose that three complementary strategy profiles be specified for G. If a solution exists (which is not taken for granted), these strategy profiles are, respectively, called a, b, and c, where the letters are indicative as follows: a is for agreement, that is, the strategy profile that the players informally agree on; b is for behaviour, the strategy profile that the players actually play post-agreement; and c is for counterfactual, that is, the strategy profile that the players would have played had they not formed the agreement a.

DSV go on to develop a specific version of their solution concept, making specific assumptions about the shape of a, b, and c. I will not describe details here because the most important part of DSV's contribution is the general principles that they propose—the idea of an underlying game and the a, b, c solution structure—rather than the specifics. Suffice it to say that DSV assumed that $a = b$, reflecting total honesty, but that the shape of a nevertheless differs from the would-be shape of a binding contract because of temptation costs that honest players have to overcome. While DSV ran an experiment and reported support for their predictions, it is clear that plausible alternative assumptions are conceivable (for example, allowing for some players to not be completely honest).

The idea of an underlying game in which the object of negotiation is a strategy profile connects DSV to some classics. With the exception of two-player zero-sum games, von Neumann & Morgenstern (1944) proceed analogously, except that they consider binding contracts (see pp. 223–224). Also, Nash (1953) considers players who strike binding contracts regarding how to play an underlying game, adding the feature that before negotiations start, the players announce 'threats' that an 'umpire' forces them to implement if they subsequently fail to reach an agreement. DSV share the outlook that strategy profiles are objects of negotiation but neither limit attention to binding contracts nor presuppose access to an umpire.

A few recent experiments indicate that the psychological impact that informal agreements have on the players who struck them may be both powerful and unique. In particular, consider the debate surrounding the studies run by Charness & Dufwenberg (2006) and by Vanberg (2008). The discussion originally focused on whether belief-based guilt aversion or a belief-independent preference for keeping one's word could best explain why communication would foster trust and cooperation. However, recent papers by Di Bartolomeo et al. (2023a, 2023b) point

out that Vanberg's communication scheme, with a format that resembles a conversation and which might plausibly generate an informal agreement, differs crucially from that of Charness & Dufwenberg, which rather involves one-sided messages likely to generate promises that cannot be requited. Di Bartolomeo et al. present results that are suggestive that the psychology involved is rather different in the two cases, which may help reconcile findings. See also Dufwenberg et al. (2023), who present results that indicate, in a field setting, that an informal agreement may be extremely powerful.

References

Baker, George, Robert Gibbons & Kevin Murphy (2002), "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics* 117, 39–84.

Ben-Ner, Avner & Louis Putterman (2009), "Trust, Communication and Contracts: An Experiment," *Journal of Economic Behavior & Organization* 70, 106–121.

Catonini, Emiliano (2021), "Self-enforcing Agreements and Forward Induction Reasoning," *Review of Economic Studies* 88, 610–642.

Charness, Gary & Martin Dufwenberg (2006), "Promises & Partnership," *Econometrica* 74, 1579–1601.

Di Bartolomeo, Giovanni, Martin Dufwenberg & Stefano Papa (2023), "Promises & Partner-Switch," *Journal of the Economic Science Association* 9, 77–89.

Di Bartolomeo, Giovanni, Martin Dufwenberg, Stefano Papa & Francesco Passarelli (2023), "Promises or Agreements? Moral Commitments in Bilateral Communication," *Economics Letters* 222, 110931.

Dufwenberg, Martin, Paul Feldman, Maroš Servátka, Jorge Tarraso & Radovan Vadovič (2023), "Honesty in the City," *Games & Economic Behavior* 139, 15–25.

Dufwenberg, Martin, Maroš Servátka & Radovan Vadovič (2017), "Honesty & Informal Agreements," *Games & Economic Behavior* 102, 269–85.

Irlenbusch, Bernd (2004), "Relying on a Man's Word? An Experimental Study on Non-Binding Contracts," *International Review of Law & Economics* 24, 299–332.

Jang, Dooseok, Amrish Patel & Martin Dufwenberg (2018), "Agreements with Reciprocity: Co-Financing & MOUs," *Games & Economic Behavior* 111, 85–99.

Kessler, Judd & Stephen Leider (2012), "Norms & Contracting," *Management Science* 58, 62–77.

Krupka, Erin, Stephen Leider & Ming Jiang (2017), "A Meeting of the Minds: Informal Agreements and Social Norms," *Management Science* 63, 1708–1729.

Levin, Jonathan (2003), "Relational Incentive Contracts," *American Economic Review* 93, 835–57.

Malhotra, Deepak & Keith Murnighan (2002), "The Effects of Contracts on Interpersonal Trust," *Administrative Science Quarterly* 47, 534–559.

McCutcheon, Barbara (1997), "Do Meetings in Smoke-Filled Rooms Facilitate Collusion?" *Journal of Political Economy* 105, 330–350.

Miettinen, Topi (2013), "Promises and Conventions – An Approach to Pre-play Agreements," *Games & Economic Behavior* 80, 68–84.

Nash, John (1953), "Two-person Cooperative Games," *Econometrica* 21, 128–140.

Vanberg, Christoph (2008), "Why Do People Keep Their Promises? An Experimental Test of Two Explanations," *Econometrica* 76, 1467–1480.

van Damme, Eric (1987), *Stability and Perfection of Nash Equilibrium*, Springer-Verlag, Berlin.

Von Neumann, John & Oskar Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton University Press.