Linguist and Philos (2014) 37:257–261
DOI 10.1007/s10988-014-9151-y

RESPONSE

Rationality in game-theoretic pragmatics

A Response to Franke (2014)

Sascia Pavan

Published online: 15 July 2014

I am very grateful to Franke (2014) for giving me the opportunity of clarifying my views. I also appreciate his epistemic justification of the solution concept I employed in my article (2013).

I have argued that iterated admissibility (or IA) can easily explain scalar implicatures. In order to respond to Franke's objections I need to embed that account in a larger strategy, to the effect that forward induction (FI) is the most appropriate solution concept in game-theoretic pragmatics.

Forward induction first appeared in the '80s in the work of Elon Kohlberg and Jean-François Mertens. It is not a sharply defined notion, it rather embodies some intuitive principles that have been formalised in several different ways. Its intuitive significance can be expressed as the principle that, when a player observes an unexpected action on the part of some opponent, whenever this is possible, she should revise her beliefs in a way compatible with the assumption that the opponent is acting rationally. It covers several formal criteria such as the *intuitive criterion* (Cho and Kreps 1987), *never a weak best response* and *stability* (Kohlberg and Mertens 1986). Also IA is a possible formalisation of FI, even if it is one of the oldest solution concepts in game theory, dating back to the '50s.

The connection between FI and IA hinges on the fact that a reasonable way of rationalising an opponent's action is to presume that he did not choose a dominated strategy. Like its very name suggests, IA is an iterative procedure, but so are many FI-based criteria, e.g. never a weak best response and some varieties of the intuitive criterion.

According to the dominant view in game theory, the acceptable solutions must be Nash equilibria. It is well known, though, that many Nash equilibria are implausible,

S. Pavan (✉)
Pordenone, Italy
e-mail: sascia.pavan@gmail.com

this sparked the search of *refinements* of equilibrium. Forward induction is one of the most important and successful of such refinements, especially when signalling games are concerned.

Grice's hypothesis is the claim that implicatures are a consequence of the fact that speakers are rational and try to cooperate. If we wish to explore its consequences with the tools provided by game theory, FI is one of the most natural solution concepts. In other terms it is just a conservative choice, in an epistemological sense (Quine and Ullian 1978, pp. 66–68).

1 Division of pragmatic labor

In 2013 I argued that IA can easily explain scalar implicatures, but Franke is right when he points out that it cannot account for the division of pragmatic labor. I would even add that no canonical solution concept can provide a complete account of it, but that the intuitive criterion (or IC) can explain one of the two halves of this phenomenon. This corroborates the present claim that FI is the right solution concept in game theoretic pragmatics.

The division of pragmatic labor is the conjunction of the following two rules:

(i) unmarked expressions receive stereotypical interpretations;
(ii) marked expressions receive non-stereotypical interpretations.

Franke considers the following example.

(1) a. Black Bart killed the sheriff.
b. $\rightsquigarrow$ Black Bart killed the sheriff in a stereotypical way.

(2) a. Black Bart caused the sheriff to die.
b. $\rightsquigarrow$ Black Bart killed the sheriff in a non-stereotypical way.

The intuitive criterion shows that rule (ii) is just a necessary consequence of (i). This applies to Franke's model and it remains valid if outside options are added. In both cases IC eliminates all the degenerate equilibria where 1b sends 1a and 2b does not send 2a. This indeed shows that outside options do not have a crucial role in the explanation of the division of pragmatic labor.

The choice of IA for scalar implicatures and of IC for the division of pragmatic labor is just a technical detail. Some more powerful criteria work in both cases—e.g. never a weak best response or stability—but they would make the exposition more complex.

The main rivals to the present account are those theories—like Franke's (2009, 2011) or Jäger (2011, 2014)—that hinge on the notion of iterated best response (or IBR). Its supporters can claim that it can account for both halves of the division of pragmatic labor in addition to scalar implicatures, so, one might ask, why not embrace this solution concept which is more successful and discard FI? There are several reasons for not doing so. In order to make this point, I can only consider Franke's work here, but I take the general claims to apply to Jäger's theory as well.

A purely game-theoretic account of the division of pragmatic labor must show than only the "Horn" equilibrium where 1b and 2b send 1a and 2a, respectively, is compatible with the principles of rationality. But there are other equilibria which satisfy the

strongest canonical solution concepts, in particular the anti-Horn equilibrium where 1b sends 2a and 1a sends 2b, already mentioned by Franke. This equilibrium is *strict*, since any unilateral deviation would yield a strictly inferior payoff to both players. According to a well-established view, a strict equilibrium is always *self-enforcing*, i.e. when it is recommended to the players 'no player has an incentive to deviate from it' (van Damme 1989, p. 476). Therefore, since 'strict equilibria possess all nice properties one can hope for' (van Damme 1991, p. 25), any supposed game-theoretic solution concept that rules out a strict equilibrium is suspicious.

A more traditional explanation of the division of pragmatic labor would presumably hinge on the *focal point effect*. It is well known to practitioners of game theory that, in games of pure coordination with several strict equilibria, the players are able to converge on one of these because the relevant strategies are salient thanks to the way they are described or presented to the players (Myerson 1991, Sect. 3.5; Schelling 1980, p. 56). Since the way a strategy is described is not part of the mathematical structure of the game, this phenomenon cannot be accounted for in purely game-theoretic terms. It is not possible to examine in detail how this applies to the present case, but the key assumption sounds plausible. It is the claim that stereotypical instances of classes of events, actions, people, or whatever, can act as focal points.

Moreover, Franke shows that IBR needs several adjustments in order to get the correct results, and I will focus on two of them. First, if the probability values assigned to sender types are taken to represent the receiver's prior probabilities, then, in some cases, IBR leads to incorrect predictions relative to scalar implicatures. In the paper, I showed that this is true of the earlier version of Jäger's theory. Franke avoids this problem assuming that the priors in the model are flat or nearly flat (2009, pp. 75–76, 132, 2011, pp. 34–41). Second, the starting point of the sequence that characterises the IBR procedure gives equal weight to all 'truthful sender strategies', on the grounds of the principle of insufficient reason (2009, pp. 54–55). No matter how plausible this principle can be in a given context, it is not valid in probability theory. Both of these additional assumptions are somewhat *ad hoc*.

Finally, the role of IBR in the game-theoretic literature is quite limited. It can explain some experimental results where people's behaviour deviates from the predictions of standard solution concepts, for example in games called 'p-beauty contests'. Yet, there is little evidence that IBR works as a general-purpose solution concept, leading to correct results in a wide class of games. Supplemented with the above additional assumptions, the scope of IBR is even more limited. As a matter of fact Franke has defined a solution concept which, as far as we know, can be applied only within the restricted realm of game-theoretic pragmatics, and therefore lacks in generality (Quine and Ullian 1978, pp. 73–78).

For these reasons, I reject as misdirected a claim toward the end of his paper when he tries to retort one of my own remarks. If our aim is to formalise Grice's hypothesis, given fairly standard principles of epistemology—conservatism and generality—there is a premium on solution concepts which have a wide applicability also outside pragmatics, even if they cannot explain all pragmatic phenomena.

2 Deviations from truth and literal meaning

My models are built in such a way that the action space of the sender does not include the possibility of saying something which is not true. In other terms, linguistic conventions and sincerity are embedded in the very structure of the game.

If this condition were relaxed, then the fact that almost always human communication conforms to linguistic conventions should be considered to be due to the near-platitude that conventions are formidable focal points (Lewis 1969). Franke is undoubtedly right when he remarks that deviations from literal meaning do not violate Grice's principle of cooperation. From Grice's principle one can at most derive that the players' interests are to some extent aligned. From this in turn, when a speaker says something which is manifestly untrue, one can infer that he is not conforming to the usual conventions, but is exploiting the literal meaning of his utterance as a creative means to direct the hearer's attention toward something else. This is arguably what happens when people use metaphors and hyperboles (Davidson 1984). Of course, this is not even a sketch of a game-theoretic account of these figures of rhetoric. I only want to show that the present theory does not preclude the possibility of such an account, and therefore that, on this respect, it is on a par with theories based on IBR, since, as far as I know, there is no IBR-based explanation of these phenomena yet.

3 Outside options

Franke complains that the introduction of outside options in the sender strategy space can lead to counterintuitive results. He considers the following two sentences.

(3) I consider it possible that p.

(4) I do not know whether p.

The point is that (4) states explicitly the full pragmatic content that (3) conveys only by implicature, yet the former does not seem to be more complex than the latter. As a consequence, any receiver response to (3) is compatible with IA.

This example is extremely interesting and it raises some deep issues that cannot be accounted for here. Let me just briefly sketch what I take to be going on. A sentence like (3) should be first contrasted with

(5) It's possible that p.

which can be justifiably asserted whenever (3) can, but is simpler. They have distinct pragmatic features, according to my intuition (3) expresses a greater uncertainty introducing an element of subjectivity which is not present in (5), presumably an instance of the division of pragmatic labor. In other terms, (3) is marked and this is an interference factor. As a consequence, instead of (4), the right outside option for the relevant sender type in the game built in order to account for the implicatures triggered by (3) is arguably a sentence whose structural and pragmatic features bear a closer resemblance to it, e.g.

(6) I consider it possible that p, but not certain.

References

Cho, I.-K., & Kreps, D. M. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics, 102*(2), 179–221.

Davidson, D. (1984). What metaphors mean. In *Inquiries into truth and interpretation* (pp. 245–264). Oxford: Oxford University Press.

Franke, M. (2009). Signal to act. Game theory in pragmatics. Ph.D. thesis, Institute for Logic, Language and Computation, Amsterdam.

Frank, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics, 4*, 1–82.

Franke, M. (2014). On admissibility in game theoretic pragmatics. *Linguistics and Philosophy*, this volume.

Jäger, G. (2011). Game-theoretical pragmatics. In J. F. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (2nd ed., pp. 467–491). London: Elsevier.

Jäger, G. (2014). Rationalizable signaling. *Erkenntnis, 79*(4), 673–706.

Kohlberg, E., & Mertens, J.-F. (1986). On the strategic stability of equilibria. *Econometrica 54*(5), 1003–1037.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Myerson, R. B. (1991). *Game theory: Analysis of conflict*. Cambridge, MA: Harvard University Press.

Pavan, S. (2013). Scalar implicatures and iterated admissibility. *Linguistics and Philosophy, 36*(4), 261–290.

Quine, W. V., & Ullian, J. (1978). *The web of belief* (2nd ed). New York: Random House.

Schelling, T. C. (1980). *The strategy of conflict* (2nd ed.). Cambridge, MA: Harvard University Press.

van Damme, E. (1989). Stable equilibria and forward induction. *Journal of Economic Theory, 48*, 476–496.

van Damme, E. (1991). *Stability and perfection of nash equilibria*. Berlin: Springer.