*Erschienen in: Logic and the foundations of game and devision theory - LOFT 8 : 8th International Conference Amsterdam, The Netherlands, July 3 - 5, 2008; revised selected papers / Bonanno, Giacomo et al. (Hrsg.). - Berlin [u.a.] : Springer, 2010. - (Lecture Notes in Computer Science : Lecture Notes in Artificial Intelligence ; 6006). - S. 135-150. - ISBN 978-3-642-15163-7
https://dx.doi.org/10.1007/978-3-642-15164-4_7*

From Nash to Dependency Equilibria

Wolfgang Spohn[*]

Fachbereich Philosophie, Universität Konstanz, Universitätsstraße 10,
78464 Konstanz, Germany
wolfgang.spohn@uni-konstanz.de

Abstract. As is well known, Nash equilibria assume the causal independence of the decisions and the actions of the players. While the independence of the actions is constitutive of normal form games, the independence of the decisions may and should be given up. This leads to the wider and quite different notion of a dependency equilibrium; e.g., cooperation in the single-shot prisoners' dilemma is a dependency equilibrium. The paper argues this notion to be meaningful and significant and sketches some of its consequences.

1 Introduction

Game theory is now 65 years old, and it has had a breath-taking career. It has become *the* foundational theory of all economics; countless economic phenomena have found a game theoretic explanation; countless economic problems have found a game-theoretic solution. No doubt, these achievements are at least comparable to the Bourbaki program in mathematics. Indeed, its claim reaches far into all of social sciences; even here it exhibits great success, although the extension of its claim is contested. The picture of a rationally deciding social individual that game theory explicates dominates large parts of our cultural and political ideology.

In all that the notion of a Nash equilibrium is *the* foundation of game theory. Almost all theoretical efforts deal with it or build on it. Many equilibrium concepts have been invented in the meantime, but almost all lie between the narrowest, the notion of a strict Nash equilibrium, and the widest, the notion of a Nash equilibrium (see the survey diagrams in [14, pp. 335f]). The notion was and remains extremely compelling, also for me.

I have become doubtful, though. The notion rests on an assumption of which all are aware and which appears entirely obvious, namely the assumption of the causal independence of the decisions and actions of the players. This assumption is unjustified, as I will show not with the help of exotic scenarios, but through a straightforward way of reasoning. If one drops the assumption, one is automatically led to the wider notion of a dependency equilibrium, as I call it. Dependency

[*] I am most grateful for having received very valuable comments from Christoph Fehige, Christian List, Ita Sher, the editors of this volume, and two anonymous referees.

equilibria behave in a significantly different way; for instance, mutual cooperation is a dependency equilibrium in the single shot prisoners' dilemma. This indicates that the foundations of game theory might undergo dramatic changes, if that notion were taken seriously—changes that would not deny, but fully turn over the picture of a rationally deciding individual and that would thus have significant theoretical and even ideological consequences.

I shall attempt to make these claims credible in the next four sections. In Section 2 I shall briefly explain the notion of a Nash equilibrium and how the said assumption is built into it. In Section 3 I shall explain which equilibrium concept ensues when this assumption is given up. In Section 4 I shall justify why the denial of this assumption is not only not absurd, but natural and often mandatory. Section 5 concludes with a few more general and comparative remarks.

I should emphasize at the outset that this is not a formal paper. I have made some rudimentary glimpses at the formal theory of dependency equilibria in [35,37]. There is not much sense, though, in developing it further, unless its meaningfulness is clearly established, something rather obscured by formal activism. Therefore, Section 4 will be the core section of this paper where I shall try to explain its meaningfulness in as plain terms as possible. Once this is accepted, we may, and indeed should, continue elaborating the formal theory.

2 Nash Equilibria

Let us focus on two-person games in normal form. The conceptual generalization of our considerations should be obvious; the transfer to other forms of game representation, in particular to the extensive form, should be thought through. Let us call the two players Ann and Bob. Ann has a set $A = \{a_1, \ldots, a_m\}$ of options; these may be few simple actions as in scissors, paper, stone or many complex strategies as in chess that prescribe a response to each possible course of the game. Correspondingly, Bob has the set $B = \{b_1, \ldots, b_n\}$ of options. The actual complete course of the game, the outcome, depends not only on their decisions, but possibly on further contingencies not fully controlled by the players, the so-called moves of nature. The outcomes or complete courses of the game are evaluated by the players in similar or different ways. In the normal form, however, Ann's and Bob's evaluations get reduced to evaluations of their possible strategy combinations that already contain their expectations with respect to the more or less favorable outcomes ensuing from the strategy combinations. So, let u be the evaluation or utility function of Ann and v that of Bob; both are functions from $A \times B$, the set of strategy combinations, into $\mathbb{R}$, the set of reals.

According to the standard theory, Ann and Bob do not only have pure, but also mixed strategies. Let S be the set of Ann's mixed strategies, i.e., the set of all probability distribution s over A; similarly, T is to be the set of all distributions t over Bob's set B of pure strategies. So, if the mixed strategy combination $\langle s, t\rangle$ is played, Ann's expected utility is

$$\sum_{i=1}^{m}\sum_{j=1}^{n} s(a_i)\cdot t(b_j)\cdot u(a_i,b_j)$$

and likewise for Bob.

Now, as is well known, a strategy combination $\langle s,t\rangle$ is a *Nash equilibrium* if and only if no player can improve by individually deviating from the equilibrium, i.e., if for all of Ann's mixed strategies $s' \in S$,

$$\sum_{i,j} s(a_i)\cdot t(b_j)\cdot u(a_i,b_j) \geq \sum_{i,j} s'(a_i)\cdot t(b_j)\cdot u(a_i,b_j)$$

or, equivalently, for all $a_k \in A$

$$\sum_{i,j} s(a_i)\cdot t(b_j)\cdot u(a_i,b_j) \geq \sum_{j} t(b_j)\cdot u(a_k,b_j)$$

and if the corresponding inequalities hold for Bob. Each game has at least one Nash equilibrium in mixed strategies. $\langle s,t\rangle$ is a *strict Nash equilibrium* iff each player can only lose by individually deviating, i.e., if "$\geq$" is replaced by "$>$" in the above inequalities. Strict Nash equilibria need not exist; but if they do, they do so only in pure strategies (where $s(a_i)=1$ and $t(b_j)=1$ for some i and j).

Traditionally, game theorists assumed that players actually choose mixed strategies by employing some random device determining the pure strategy to be played. In epistemic game theory, grounding in many papers of John Harsanyi, more explicitly conceived, e.g., in [1] and [33], and finally established in the theory of rationalizability of [7] and [27], a different interpretation emerged that is more appropriate for our context: A Nash equilibrium $\langle s,t\rangle$ may also be conceived as an *equilibrium of opinions*. For, why should Ann choose mixed strategy s[1]? This can only be reasonable when she does not care which of the pure strategies a_i with $s(a_i)>0$ results from playing s. But how can she be indifferent? Only when all a_i with $s(a_i)>0$ are equally good for her, i.e., have the same expected utility $\sum_j t(b_j)\cdot u(a_i,b_j)$—where t now represents Ann's opinion about Bob's pure strategies. This indifference is guaranteed in the equilibrium $\langle s,t\rangle$. The same holds for Bob's pure strategies b_j with $t(b_j)>0$, when s represents his opinion about Ann's possible actions. Only in such an equilibrium of opinions can the opinions of the players be mutual or common knowledge among the players (as it has been assumed in game theory all along with respect to the structure of the game and the utility functions of the players)[2]. Ann cannot stick to her opinion t about Bob and at the same time guess that Bob may not

[1] Cf. also the old critical considerations of [13].

[2] Common knowledge usually denotes the full infinite hierarchy of mutual (and reflexive) beliefs. As observed in [33, sect. 4] and emphasized in [4] first-order and mutual second-order beliefs suffice, at least in the two-person case; Aumann & Brandenburger ([4]) call this mutual knowledge. For this paper it is not important to carefully distinguish between mutual and common knowledge (and, for that matter, between knowledge and belief).

have the opinion s about her (as long as she is convinced that he is an expected utility maximizer).

Now one may further ponder about the justification of such equilibria. One may derive the rationality of the opinion equilibrium from the rationality of the mixed strategy equilibrium; if it is rational for Ann to play the mixed equilibrium strategy s and if Bob takes Ann to be rational, then Bob must rationally have the opinion s about Ann. This is how one can view the matter provided one has already justified the rationality of the mixed strategy equilibrium in some other way. If one prefers to do without that presupposition—as I do—, one may try to directly show the rationality of the opinion equilibrium or to derive if from common knowledge or common prior assumptions. I have extensively discussed all this in [33], with the somewhat skeptical conclusion that justification carries only up to the rationalizable strategies, as they were called and more deeply investigated by Bernheim ([7]) and Pearce ([27]). The issue is further elaborated by Aumann & Brandenburger ([4]), in particular for the more complicated more-than-two-person case.

Be this as it may, let us accept the familiar assumption that what is rational must be sustainable in public, i.e., may be mutual or common knowledge. Then, obviously, only Nash equilibria can be maintained as rational. However, this is the gist of the matter, the crucial assumption indicated in the introduction is already contained in this abstract representation of the social situation of Ann and Bob. Let me explain:

In a mixed strategy equilibrium $\langle s, t\rangle$ Ann and Bob play their strategies each by their own; there is no playing of a joint distribution p over $A \times B$, as it is the case with Aumann's [2,3] objective correlated equilibria. The tossing of the one has no causal influence on the tossing of the other; s and t are assumed to be causally independent. Of course, this holds as well for the actions or pure strategies selected; what Ann does has no influence on what Bob does, and vice versa.[3]

Well, how could Ann's action have an influence on Bob's? The usual way would be that Bob sees, or is in some other way informed about, what Ann does and behaves accordingly. However, this is explicitly excluded; it would have to be modeled in a different way. Of course, subconscious influences or even more exotic scenarios are excluded all the more. The same holds for the preferred weaker interpretation of $\langle s, t\rangle$ as an opinion equilibrium. Ann's opinion consists in an unconditional distribution t over Bob's possible actions which she thereby accepts as something she cannot influence, just as, say, tomorrow's weather. If she thought to have at least probabilistic influence on Bob's actions, then her probabilities for Bob's actions should depend on her own actions, i.e., unlike t

[3] At this point it is advisable to conceive pure strategies as single actions. The latter clearly stand in causal relations. Whether one can say so of complex contingency plans is at least questionable, since they are temporarily and modally extended, i.e., they plan for many possible situations most of which do not realize. In fact, "plan" already contains the ambiguity I shall emphasize later on: the adoption of a plan may be a local event capable of causal relations, its execution is not.

they should vary with her own actions. Vice versa for Bob. That is, if we conceive the opinion equilibrium as a Nash equilibrium, we have built into it the belief of the players in the mutual causal independence of their actions.[4]

To be sure, the last conclusion is not entirely cogent. It presupposes a relation between conditional subjective probabilities and causal opinions that is in need of justification; after all, deterministic and probabilistic causation are most contested notions. However, the relation as I have just presupposed it appears most plausible and has strong backings.[5] I shall return to it in Section 4.

This is the snag. I shall certainly not doubt the causal independence of Ann's and Bob's actions; that would be absurd. The causal independence of Ann's and Bob's decisions or intentions, however, is something subtly different. It is the point I shall question; and this will then have specific consequences for the form of Ann's and Bob's subjective probabilities.

In Section 4 I shall explain what this may mean. For the moment, I only want to dismiss the conclusion that Ann's and Bob's opinions about each other consist in unconditional subjective probabilities. If we give this up, the concept of a Nash equilibrium is no longer applicable. What could stand in its place? This is the topic of the next section.

3 Dependency Equilibria

What we have to do now is to allow that Ann's conditional probabilities for Bob's actions or pure strategies vary. Thus, her opinions now take the form $q(b_j \mid a_i)$ where for each $a_i \in A$, $q(\cdot \mid a_i)$ is a probability distribution over B. Reversely, Bob's opinions now take the form $r(a_i \mid b_j)$, where for each $b_j \in B$, $r(\cdot \mid b_j)$ is a distribution over A.

What could it mean under these assumptions for Ann to be rational? It means to *maximize conditional expected utility*, i.e., to choose an a_i for which $\sum_i q(b_j \mid a_i) \cdot u(a_i, b_j)$ is maximal. This was the important progress of Fishburn ([17]) over Savage ([30]). Savage assumed the subject to have only unconditional probabilities for action-independent circumstances or states of the world. Fishburn found Savage's distinction of states of the world and consequences and the accompanying conception of acts as functions from the former to the latter to be problematic and to be made superfluous by his assumption that the subject has probabilities conditional on each of her possible acts for the rest of world (where the acts are a primitive ingredient of the decision model); of course, this allows that some propositions, e.g., Savage's states of the world, are probabilistically independent of the actions. In this conception, the subject then maximizes conditional expected utility. This is, I believe, generally accepted.

So far, Ann's and Bob's subjective probabilities were not constrained by any conditions. However, we shall now take over the leading idea entailing Nash

[4] Aumann & and Brandenburger ([4]) outright assume the opinions or conjectures about the other players to be of that unconditional form entailing causal independence. Hence, their results do not constrain my further considerations.

[5] Cf. [32, sect. 3.3] [31, Sect. 3.7], [25], and [28, Ch. 3 and 4].

equilibria and assume that these probabilities are no secrecy, but must be sustainable in public, i.e., can be or even are common knowledge. This entails two substantial constraints.

The first constraint did not show up in Nash equilibria, since they satisfied it, anyway. It is that Ann's and Bob's conditional probabilities must combine into a single joint distribution p over $A \times B$, i.e., there must be such a p such that for all i and j, $p(b_j \mid a_i) = q(b_j \mid a_i)$ and $p(a_i \mid b_j) = r(a_i \mid b_j)$. This condition may fail to be satisfied, since q and r together have more degrees of freedom than p.

The combinability of Ann's q and Bob's r into a joint distribution p follows, however, from the possibility of common knowledge. If such a joint distribution p does not exist, then Ann cannot know Bob's probabilities r, moreover know that Bob knows her probabilities q and still stick to her q. This is so far just a consistency constraint that, as mentioned, was automatically satisfied in the case of Nash equilibria and that has the consequence that from now on we may restrict attention to that joint distribution p.

The second constraint is induced by the common knowledge of rationality. According to the p in question we have $p(a_i) > 0$ for some of Ann's actions $a_i \in A$; hence, each such a_i must have positive probability in Bob's view conditional on at least some of his actions. How can this be? Bob knows that Ann is rational, i.e., that Ann maximizes conditional expected utility. If Ann achieves this only with a_i and no other action, we should have $p(a_i) = 1$ and Bob should be certain that Ann does a_i. If, however, several of Ann's actions are optimal in Ann's view and Bob rightly assigns them positive probability, then all of them must have equal and maximal conditional expected utility for Ann. Mutatis mutandis, the same holds for Bob. (Certainly, the derivation of the two constraints should be carried out in formal detail.)

Thus, the mutual or common knowledge or knowability assumption leads us to the following equilibrium concept. The probability distribution p over $A \times B$ is a *dependency equilibrium* iff for all i with $p(a_i) > 0$ and all $k = 1, \dots, m$

$$\sum_j p(b_j \mid a_i) \cdot u(a_i, b_j) \geq \sum_j p(b_j \mid a_k) \cdot u(a_k, b_j)$$

and reversely, for all j with $p(b_j) > 0$ and all $l = 1, \dots, n$

$$\sum_i p(a_i \mid b_j) \cdot v(a_i, b_j) \geq \sum_i p(a_i \mid b_l) \cdot v(a_i, b_l)$$

i.e., if all of Ann's and Bob's actions that are not excluded and have positive probability according to p have, respectively, the same maximal expected utility for Ann and Bob.

There is no space and no need to formally develop this concept here. I restrict myself to a few remarks. First, for some $a_k \in A$ or $b_l \in B$ we may have $p(a_k) = 0$ or $p(b_l) = 0$ so that no conditional probabilities are defined for them and the definition just given makes no sense. This defect may, however, be removed in a precise and adequate way (cf. [37, Sect. 2]).

Second, I should remark that dependency equilibria are not to be confused with the correlated equilibria of [2,3]. An objective correlated equilibrium is also a joint distribution p over $A \times B$. However, the issue here is roughly whether or not it pays for a player to break the dependency given by p and do something instead that is optimal under the marginal distribution over the actions of the other player given by p. If there is no such advantage, then no player will break the dependency and p is a correlated equilibrium. Even this very coarse description shows that correlated and dependency equilibria are driven by different ideas.

Dependency equilibria form a wider class than Nash equilibria. Those distributions p over $A \times B$ that factorize into independent s over A and t over B—Nash equilibria apply only to such p—are obviously Nash equilibria if and only if they are (degenerated) dependency equilibria according to our definition. The way in which dependency equilibria go beyond Nash equilibria (and also diverge from correlated equilibria) is illustrated in [37, Sect. 3] with some significant examples.

The most important of them is the single-shot prisoners' dilemma. Its normal form is given, for instance, by the utility matrix of Table 1, where Ann is the row chooser and Bob the column chooser and where c stands for "cooperate" and d for "defect". It is obvious that d strictly dominates c and is thus preferred to c, given the independence of the other's choice. Hence, $\langle d, d\rangle$ is the only Nash equilibrium and even the only correlated equilibrium in the prisoners' dilemma.

Table 1. The prisoners' dilemma. For a pair (u, v), u represents the utility of Ann, and v that of Bob.

	c	d
c	(2,2)	(0,3)
d	(3,0)	(1,1)

However, there surprisingly are two whole families of dependency equilibria, one in which the players are asymmetrically or negatively correlated and one in which they are symmetrically or positively correlated. The latter are given by the matrix of Table 2 for all x with $0 \leq x \leq 1$.

Table 2. Two families of dependency equilibria for the prisoner's dilemma

p	c	d
c	$\frac{1}{2}x(1+x)$	$\frac{1}{2}x(1-x)$
d	$\frac{1}{2}x(1-x)$	$\frac{1}{2}(1-x)(2-x)$

The fact that both, cooperation and defection, have the same conditional expected utility according to such a p must be checked by calculation. It is, however, immediately obvious that $p(\langle c, c\rangle)$ converges to 1 if x does and that $p(\langle c, c\rangle) = 1$ for $x = 1$. That is, $p(\langle c, c\rangle) = 1$ or certain mutual cooperation is a dependency equilibrium that consists in each player's belief that the other will cooperate if and only if he or she cooperates. It is even the weakly Pareto-dominant dependency equilibrium; in no other dependency equilibrium do the players have a higher conditional expected utility.[6] For details see again [37, Sect. 3]. I shall return to the example.

I admit, however, that so far the theory of dependency equilibria is developed only in a most rudimentary way. This is due to their novelty and their perhaps doubtful significance, but also due to mathematical intricacies (in n-person games one has to solve systems of quadratic equations for many variables instead of linear equations). At least, though, we have in the two-person case that each pure strategy combination $\langle a_i, b_j \rangle$ or each p with $p(a_i, b_j) = 1$ that weakly Pareto dominates a Nash equilibrium is a dependency equilibrium (for a proof see [35, pp. 208f.]). Ita Sher pointed out to me that the same sort of proof should apply for showing that exactly those pure strategy combinations are dependency equilibria that weakly Pareto dominate the maximin strategies of the players. That suggests at least that exactly those combinations of pure strategies that are at least as good as the maximin strategies and that are not Pareto dominated by other strategy combinations are the most interesting from the point of view of the theory of dependency equilibria.

Since Nash equilibria are also dependency equilibria, the existence of the latter is guaranteed. However, the selection problem is graver for the latter. With respect to Nash equilibria this problem was that even in view of many Nash equilibria one may hope to be able to justify a rational selection from them and thus to come to more specific recommendations. Whether this is feasible is contested within the standard theory.[7] As for dependency equilibria, I have presently nothing to say on this issue.

Still, already these few remarks suggest that game theory might considerably change when based on the notion of dependency instead of Nash equilibria. This brings us back to the question: Why should we take this notion seriously? So far, it must be understood as imputing to both players the belief that their actions have a causal influence on the other's action. Since such causal loops are impossible—it would be bizarre to deny this—, at least one of the players and presumably both must be massively in error. However, a notion that builds on such error is uninteresting; it could not be common knowledge.

At the end of Section 2 I had claimed that this is not our problem. This claim is still very mysterious. The next section attempts to solve the mystery.

4 The Causation of and Correlation between Actions

The argument I shall give now may sound involved; its core, however, is quite simple. In [35, sect. 3-5] I have elaborated a formal version. Here, I venture an

[6] This consideration seems to resemble the justification of mutual cooperation by the so-called mirror principle (cf. [16]). However, this justification always seemed to me to be inconclusive ([35, p. 250]) and to receive sufficient support only through dependency equilibria. Moreover, the mirror principle works only when the situation of the players is perfectly symmetric, whereas the theory of dependency equilibria is not restricted in this way.

[7] [20] certainly was the most heroic attempt at a general solution of the selection problem.

informal short version that should be much more perspicuous, even though it does not render the formal version superfluous.

As stated, our problem is to find an explanation of the action-dependent probabilities for the other player's actions assumed by dependency equilibria, an explanation that does not impute mad causal beliefs to the players.

The first thought may be that there is no problem at all; this is simply the old story of cause and correlation. Of course, any two variables and thus also two variables like the actions of the two players may be correlated, i.e., probabilistically dependent without being causally dependent, i.e., without the earlier exerting causal influence on the later. The most we can assert is Reichenbach's principle of the common cause according to which two correlated variables must have a common cause, indeed a complete common cause rendering them independent, if neither causally depends on the other.[8]

In general, this is correct, no doubt, and I shall return to it. However, the second thought must be that this general point does not apply at all from the point of view of an agent, as far as his own actions are concerned. From his point of view, i.e., within his model of the decision situation, his actions are exogenous variables that have only effects, but no causes within the model. The rational agent wants to optimize the probable consequences of his actions; causes of his actions, whether they consist in his practical deliberation itself or lie elsewhere, simply play no role in the optimization of consequences.[9] Hence, the agent cannot view a correlation between his actions and other variables as indicative of a common cause. In this special case, the correlation always represents a causal dependency.

This insight is the whole point of the long-standing discussion about Newcomb's problem (cf. e.g., [12]), in which the great majority has adopted the stance of causal decision theory that is characterized by this insight. If, counterfactually, my desire to smoke as well as my increased risk of lung cancer would exclusively be genetically caused and thus correlated, it would be silly to let the correlation ruin my desire; by refraining to smoke I cannot reduce my risk of lung cancer. The exogeneity of the action variables has been theoretically developed in the theory of truncated graphs ([28, sect. 3.2]) and of manipulated graphs ([31, sect. 3.7]); but see already [32, sect. 3.3 and 5.2]). Intuitively, though, the point was clear long before all those theoretical discussions. As a consequence, the only possibility to conceive equilibria was to conceive them as Nash equilibria.

In general, also this second thought is correct. However, it has a gap in turn: When the deciding of the agent, i.e., his decision situation (that is nothing but his view of it), causally influences, in his view, not only his action, but also

[8] The principle is widely, though not universally accepted; it is, for instance, a consequence of the Causal Markov Condition, the basic causal axiom of [31, sect. 3.4.1]. In my view it is even provable on the basis of a suitable explication of probabilistic causation; cf. [34].

[9] I think that the uncausedness of one's own actions from the perspective of one's own practical deliberation is a fundamental sense of freedom of action and of the will; cf. [36].

some other variable X, then and only then it is the case that his action, though not causally influencing the variable X in his view, is nevertheless correlated with the variable X, *in a way relevant to his practical deliberation*, i.e., to the conditional expected utility of his actions. That's the crux I want to explain.

Apparently, it is now somehow important that the actions are correlated with other variables due to common cause relations. Hence, we must first ask how actions are caused at all. Obviously in most complex ways; the most multifarious circumstances have an influence on our doings. However, since we consider only rational action by rational agents, all the influences must channel through the beliefs and the desires of the agent, through his representation of the decision situation including his subjective probabilities and utilities that result in an intention or decision for a certain action. This decision situation is the direct cause of that action according to the causal theory of agency of Hempel ([22]) and Davidson ([15]) that is still the dominating one today. (Of course, there must also exist some opaque mediation from mental state to bodily movement, something we need not consider here.)

Each decision situation causes at least one action, indeed an action optimal according to it. It might also cause several actions; possibly, a whole course of action comes to be decided. Reversely, in the view of the agent each action can be caused only by exactly one decision situation: at least one, since otherwise it would not have been an intentional action, and at most one, because one cannot think to decide twice about one and the same action; if the agent envisages to take a second decision on the same action later on, he cannot think to make a decision at the first time.

Moreover, a decision situation is a complete cause of the actions decided in it; there are no other causes of them besides the ingredients of the situation. Finally, the decision situation is only causally, but not necessarily temporally directly before the action; the decision is not always taken in the last moment.

No doubt, this is a somewhat crude, idealized picture of the relation between decision situations, i.e., complex sets of graded beliefs and desires, and actions. I think, though, that what we can say about this idealized picture holds as well for more realistic, softer pictures. In particular I should emphasize that what I have called a decision situation in which a decision is made about the possible actions or options is nothing very well or sharply defined. The conscious introspection of one's desires and beliefs and the explicit practical deliberation that results in a determinately localizable formation of an intention or decision is rather seldom; to carry out this exercise all the time would be much too demanding. Otherwise, however, it is often not so clear what the relevant beliefs and desires are and when exactly an intention has been formed.

This is not to say, however, that the intention is never really formed or only in the last moment, when the action is about to be executed and can no longer be stopped. It is important to realize that in principle decision and action are temporally decoupled. Often I go to bed with a plan in mind that I simply execute the next morning. In our strange times, where to make the next vacation is something to be decided many months before. And I lay down my last will in the hope it will take 30 years or more till it comes into force.[10]

The crucial point now is this: If we thus explicitly consider actions as caused by decision situations, then we must as well allow that such complexes of beliefs, desires, and intentions have other effects besides the relevant actions and hence are common causes of actions and other things. We sometimes declare our intentions (even though this should perhaps be modeled as a further action). Often, our intentions, being closely connected to our emotions, can be glanced from our mimics, gestures, and other emotional responses. This is most important for human intercourse.[11] Extremely controlled persons who allow a glimpse into their inner life only through their actions are somehow eerie. The much discussed toxin puzzle (see [23]) focuses exactly on this point by its fictitious story where only the forming of an intention (as measured by a cerebroscope) and not the intentional action itself is rewarded.

How should the agent deal with such a possibility? Should he take into account such side effects of his decision? Yes, by all means. If in the toxin case the forming of an intention is rewarded, then I form the intention, provided the reward outweighs the negative consequences of carrying out the intention.[12] If in the smoker's case neither the relevant genes nor the smoking itself, but, in some mysterious ways, only the desire to smoke disposes to lung cancer, then I better do not have the desire (and consequently do not smoke).

Yet, how can this point be accounted for in the representation of a decision situation? So far, this representation only contains the possible actions, all the other things or propositions the beliefs and desires are about, and, of course, the beliefs and desires themselves, but it does not, as it were, reflexively contain the possible decision situation itself as an additional variable. How could it? As we said, the causes of one's own actions are not part of the practical deliberation.

At this point, things become a bit involved.[13] One can also introduce reflexive decision models and study their relation to the non-reflexive models hitherto considered. This relation results from the fact that the reflexive and the non-reflexive model represent in a sense the same decision situation (cf. [35, sect. 4.3]). The consequence of such a refined modeling should be clear, though. In the non-reflexive model the side effects of the decision situation show up precisely in a correlation between these side effects and the actions. Such a correlation finds no explanation at all in the non-reflexive model and could then, as we saw, never be regarded as admissible. In the reflexive expansion, however, we see a common cause of this correlation, namely the (non-reflexive) decision situation itself. *It is in this case and only in this case that in the perspective of the agent his own*

[10] The temporal decoupling of decision and action is also an essential ingredient of Bratman's account of intention, planning, and agency; see [11] and [10, chs. 1-4].

[11] Frank ([18]) profoundly elaborates on this aspect of rational action.

[12] I have argued this in detail in [35, sect. 5.1].

[13] One may even resort to Barwise ([5]) who devised model theoretic means based on set theory without the foundation axiom in order to accommodate such reflexive phenomena. I think, though, that such a move is not required.

actions can be correlated with other things or variables without (probabilistically) causally influencing them.

In this way, my line of argument returns to the dependency equilibria. For, we have seen now that the correlation between the actions of the players stated in these equilibria need not be interpreted as the false belief of causally influencing the other's actions by one's own actions. It can instead point to a common cause in the way indicated.

What is the common cause in this case? If each player's action is caused exclusively by her or his decision situation, and if this decision situation is moreover to exert influence on the other player's action, then the common cause can only consist in the joint formation of the player's decision situations. Is this plausible? Yes, I think so. I had mentioned already that such a decision situation is not sharply defined. It is temporally extended, unfolding in this extension; this leaves enough room for interactions and mutual dependencies that may arise from any form of communication between the players. In particular, it seems entirely plausible to me in the original prisoners' dilemma story that the two gangsters do not take their decisions in the separate prison cells, as maliciously suggested by the police, but are decided all along, entangled in the cooperative dependency equilibrium that has formed in their sworn community during their raids.

Let me summarize once more the picture that has thus emerged with respect to the prisoners' dilemma. The players' decision situations may develop over a certain stretch of time, and they may causally interact during this time. The effect is that the one player's decision situation is causally responsible not only for her or his own action, but also for the other player's decision situation and thus indirectly for his or her resulting action. If this process evolves under conditions of mutual or common knowledge, it must result in a dependency equilibrium. That may be any, but rationally it should be the Pareto-optimal cooperative dependency equilibrium in which each player believes the other to cooperate if and only if she or he cooperates and in which cooperation maximizes conditional expected utility.

Of course, the players are free to break the mutual dependency; as stated, Nash equilibria are also dependency equilibria, though degenerated ones. The time of decision can always be chosen so late that causal interaction is excluded not only between the actions, but also between the decision situations. Sometimes, however, it is more reasonable to maintain the dependency than to break it. This is so at least in the prisoners' dilemma.

5 Afterthoughts

I have thus reached the primary aim of my paper: namely to identify the apparently indispensable assumption that committed us to Nash equilibria, to explain its dispensability and thus to provide the wider concept of dependency equilibria with sense and significance. However, one will ask, did more than 50 years of proliferous game-theoretic thinking not produce similar ideas? Yes and no.

First, I should mention that my proposal closely resembles the old theory of conjectural variation. It seems, though, that this theory was abandoned precisely for the reason that "at the level of simultaneous decisions in a single-period model, conjectural variation is not meaningful" ([19, p. 107]). So, one tried to make sense of it rather in the context of repeated games (cf. [19, sect. 5.1 and 9.3] and [29]). If I am right, this reason need not apply. Hence, it might be worthwhile elaborating on this resemblance. Conversely, of course, the theory of dependency equilibria needs to be extended to the context of repeated games, which might provide a theoretical explanation of how such a dependency may arise.

There is a growing literature on correlation within the context of evolutionary game theory (cf., e.g., [6]). In particular, there are evolutionary explanations of the emergence of cooperation in the prisoners' dilemma (cf. [21]. However, evolutionary game theory is governed by a different interpretation, an issue I cannot pursue here.[14]

Cooperative game theory does not look at how groups of players function internally, but rather presupposes correlation within possible coalitions, turning then to its own difficult problems. However, my proposal is located within non-cooperative game theory, and it might help understanding that inner functioning. For instance, it seems that any value of the characteristic function of a cooperative n-person game is the outcome of some dependency equilibrium of that game.

In non-cooperative game theory as well correlation plays an increasing role. The main strand certainly originates from Aumann's [2,3] invention of correlated equilibria. However, I had indicated how dependency equilibria diverge. The divergence continues with more recent inquiries. For instance, it is important to Brandenburger & Dekel ([8]) and Brandenburger & Friedenberg ([9]) that in the eyes of each player the acts or strategies of the other players may be correlated (in the more-than-two person case where there are at least two other players); the latter even say (p. 32) that this "is really just an adaptation to game theory of the usual idea of common-cause correlation". Even one's own acts may be correlated with those of the other players due to uncertain external circumstances on which all players' acts depend according to their strategies. Still, the optimality of strategies is assessed according to their unconditional expected utility, as is characteristic of correlated equilibria, and not according to their conditional expected utility.[15]

There also are attempts to model mutual dependency as described in the previous section within what one might call standard non-cooperative game theory. In particular, Harsanyi & Selten ([20, pp. 4-7 and 18-23]) have enlarged the prisoners' dilemma by a game of previous self-commitment moves and then showed that the Nash equilibrium of the enlarged game includes cooperation in the

[14] In [35, pp. 251ff.] I have made more extensive comparative remarks.

[15] This is true also of the a posteriori equilibria of [8, p. 1395]), which are based on a comparison of conditional expected utilities, but in a different sense; each act is compared with other acts conditional on the same information.

prisoners' dilemma. And Myerson ([26, pp. 249–257]) has generally explained how a theory of so-called preplay communication is able to reduce Aumann's correlated equilibria and the dependence encoded in them to Nash equilibria. These considerations are most instructive, and one should find out whether they might be applied to dependency equilibria as well.

However, such a move should not be interpreted as possibly reducing away dependency equilibria (or, for that matter, correlated equilibria). Of course, one may try to describe such dependencies as the result of special games within the theory of Nash equilibria. Or one may reversely consider the existence of such dependencies as given and then, as I started to do, develop a theory about which behavior is rational when standing in such dependencies and thus also about how such dependencies are rationally to be shaped. The point then is not which theory is more general; that depends on the perspective. Dependency equilibria are obviously more general than Nash equilibria; and if the ideas just mentioned were successful, one might reversely be able to represent dependency equilibria as Nash equilibria in special games. The point is rather the radical change of our conception of rationality in game situations that comes with directly considering dependency equilibria.

For instance, one usually thinks that defection would be completely rational in the one-shot prisoners' dilemma and that cooperation in the finitely iterated prisoners' dilemma can only be understood as some form of bounded rationality. It is, however, just the other way around. Cooperation is perfectly rational in the single-shot as well as in the iterated prisoners' dilemma, and defection can only be explained by insufficient trust in the rationality of the other player or by insufficient common knowledge of rationality. One must free oneself from the idea that standard decision and game theory *define* what is rational; they only *propose* an explication of rationality that has deficiencies which show up in the iterated prisoners' dilemma and elsewhere in a particularly drastic way and that must hence be improved. (And, as I would like to add, changes in one's explication of rationality also change one's conception of bounded rationality.)

Broadening the perspective, one may even say that the ideological picture propagated in particular by economic theorizing is thereby falsified, the liberalistic picture of the freely and independently deciding individual, against whose self-interest the rationality of cooperation is hard to explain. This picture turns out wrong on the basis of the present considerations. We always stand in interpersonal dependencies; and individual rationality may tell us to acknowledge these dependencies to our own benefit (and that of others) and thus to lean towards a communitarian perspective.

Genuine dependency equilibria always involve an element of commitment. I mentioned above that the time of decision can always be chosen so late as to exclude causal interaction between the decision situations. Not deferring the decision thus means remaining committed. Surely, this is another large topic for economics as well as for philosophy[16] that I should not enter. However, it seems within the reach of my considerations to provide rationality criteria for the

[16] E.g., see again Bratman ([11,10]).

choice of decision times, for early commitment or late decision, for dependence or independence and thus to integrate socalled sophisticated and resolute choice introduced and discussed by McClennen ([24]) as two competing decision rules into a unified theory.

Turning even more to philosophy, a last big topic involved in dependency equilibria is that of shared or joint or we-intentions and attitudes[17] as pursued by Raimo Tuomela for decades (see, e.g., [38]) and others (see, e.g., [10, ch 5-8]). Perhaps, dependency equilibria help accounting for the internal structure of such joint intentions, a point related to my brief remark on cooperative game theory. Again, though, this is a suggestion I cannot pursue here.

These associations open a rich agenda. Simply mentioning them is certainly bad scientific style. However, I wanted to suggest at least that the notion argued here to make good sense is indeed at the intersection of many pressing issues. Good reason to pursue it further.

References

1. Armbruster, W., Böge, W.: Bayesian Game Theory. In: Moeschlin, O., Pallaschke, D. (eds.) Game Theory and Related Topics, pp. 17–28. North-Holland, Amsterdam (1979)
2. Aumann, R.J.: Subjectivity and Correlation in Randomized Strategies. J. Math. Econ. 1, 67–96 (1974)
3. Aumann, R.J.: Correlated Equilibrium as an Expression of Bayesian Rationality. Econometrica 55, 1–18 (1987)
4. Aumann, R.J., Brandenburger, A.: Epistemic Conditions for Nash Equilibrium. Econometrica 63, 1161–1180 (1995)
5. Barwise, J.: On the Model Theory of Common Knowledge. In: Barwise, J. (ed.) The Situation in Logic. CSLI Lecture Notes, vol. 17, CSLI, Cambridge (1990)
6. Bergstrom, T.C.: The Algebra of Asortative Encounters and the Evolution of Cooperation. Int. Game Theory Rev. 5, 211–228 (2003)
7. Bernheim, B.D.: Rationalizable Strategic Behavior. Econometrica 52, 1007–1028 (1984)
8. Brandenburger, A., Dekel, E.: Rationalizability and Correlated Equilibria. Econometrica 55, 1391–1402 (1987)
9. Brandenburger, A., Friedenberg, A.: Intrinsic Coerraltion in Games. J. Econ. Theory 141, 28–67 (2008)
10. Bratman, M.E.: Faces of Intention. Selected Essays on Intention and Agency. Cambridge University Press, Cambridge (1999)
11. Bratman, M.E.: Intentions, Plans, and Practical Reasons. Harvard University Press, Cambridge (1987)
12. Campbell, R., Sowden, L. (eds.): Paradoxes of Rationality and Cooperation. University of British Columbia Press, Vancouver (1985)
13. Chernoff, H.: Rational Selection of Decision Functions. Econometrica 22, 422–443 (1954)
14. van Damme, E.: Stability and Perfection of Nash Equilibria, 2nd edn. Springer, Berlin (1991)

[17] Thanks to Christian List for pointing this out to me.

15. Davidson, D.: Actions, reasons, and causes. J. Philos. 60, 685–700 (1963)
16. Davis, L.: Prisoners, Paradox, and Rationality. Am. Philos. Q. 114, 319–327 (1977)
17. Fishburn, P.C.: Decision and Value Theory. Wiley, New York (1964)
18. Frank, R.H.: Passions Within Reason. The Strategic Role of the Emotions. W. W. Norton & Company, New York (1988)
19. Friedman, J.: Oligopoly Theory. Cambridge University Press, Cambridge (1983)
20. Harsanyi, J.C., Selten, R.: A General Theory of Equilibrium Selection in Games. MIT Press, Cambridge (1988)
21. Heiner, R., Albert, M., Schmidtchen, D.: Rational Contingent Cooperation in the One-shot Prisoner's Dilemma (2000) (unpublished manuscript)
22. Hempel, C.G.: Rational action. Proc. Addresses APA 35, 5–23 (1961)
23. Kavka, G.S.: The Toxin Puzzle. Analysis 43, 33–36 (1983)
24. McClennen, E.F.: Rationality and Dynamic Choice. Cambridge University Press, Cambridge (1990)
25. Meek, C., Glymour, C.: Conditioning and Intervening. Br. J. Philos. Sci. 45, 1001–1021 (1994)
26. Myerson, R.B.: Game Theory. Analysis of Conflict. Harvard University Press, Cambridge (1991)
27. Pearce, D.G.: Rationalizable Strategic Behavior and the Problem of Perfection. Econometrica 52, 1029–1050 (1984)
28. Pearl, J.: Causality. Models, Reasoning, and Inference. Cambridge University Press, Cambridge (2000)
29. Sabourian, H.: Rational Conjectural Equilibrium and Repeated Games. In: Dasgupta, P., Gale, D., Hart, O., Maskin, E. (eds.) Economic Analysis of Markets and Games, pp. 228–257. MIT Press, Cambridge (1992)
30. Savage, L.J.: The Foundations of Statistics. Wiley, New York (1954); 2nd edn., New York, Dover (1972)
31. Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search. Springer, Berlin (1993); 2nd edn. (2000)
32. Spohn, W.: Grundlagen der Entscheidungstheorie. Ph.D. thesis, Universität München (1976)
33. Spohn, W.: How to make sense of game theory. In: Stegmüller, W., Balzer, W., Spohn, W. (eds.) Philosophy of Economics, pp. 239–270. Springer, Berlin (1982)
34. Spohn, W.: On reichenbach's principle of the common cause. In: Salmon, W.C., Wolters, G. (eds.) Logic, Language, and the Structure of Scientific Theories, pp. 215–239. Pittsburgh University Press, Pittsburgh (1994)
35. Spohn, W.: Dependency equilibria and the causal structure of decision and game situations. Homo Oeconomicus 20, 195–255 (2003)
36. Spohn, W.: The core of free will. In: Machamer, P., Wolters, G. (eds.) Thinking About Causes. From Greek Philosophy to Modern Physics, pp. 297–309. Pittsburgh University Press, Pittsburgh (2007)
37. Spohn, W.: Dependency equilibria. Philos. Sci. 74, 775–789 (2007)
38. Tuomela, R.: Cooperation. Kluwer, Dordrecht (2000)