DEGREES OF FREEDOM IN THE SOCIAL WORLD

Mariam Thalos
Philosophy Department, SUNY Buffalo

Abstract: Ever since Hobbes we have sought to explain such extraordinarily commonplace facts as that prudent people trust each other, keep promises and succeed by and large at coordinating so as not to collide in corridors and roadways. In fact, it is the success which meets our efforts to coordinate without communication so as not to collide in corridors, which is perhaps the most perplexing of all social phenomena, because it is enjoyed by creatures with the capacity to conceive of (inconveniently) many coordination schemes that achieve the same ends. And ever since Thomas Schelling we have been suspecting that traditional game theory, which as I shall show embraces a Kantian view of agency, cannot explain this success. I shall here propose an anti-Kantian account of agency which takes as its point of departure the kinds of coordinations that Schelling demonstrated are so difficult for traditional game theory to explain. My proposal shall reject the Bayesian foundations of game theory, which rest on this Kantian view of agency.

1. Bayesian "Coordinations"

Bayesian foundations of decision theory aim at treating cooperative enterprises or "games", defined as ones which permit enforceable agreements amongst the players,[1] as they treat non-cooperative games, which admit of no such agreements. And so Bayesians aim at representing collective decision making as a species of individual decision making, not as something different in kind. The goal is to reduce what may be called "collective rationality", the process of deliberating collectively to achieve agreement for the sake of coordinating action, to what may be called "individual rationality," the process of achieving decision as individuals for whom overt deliberation is a (bargaining) game of individual-against-individual, in its own right, within the boundaries of the larger game. The idea is to handle the overt process of deliberation as a series of strategic bargaining moves in a competitive game of individual-against-individual, which is also governed by rules specified in advance, and therefore *not* as the sort of thing which someone can undergo purely as a single individual. Rather than being seen as a potential means to reaching collective decision with others, and therefore as a means of forming a single but larger decision-making body aiming at common goals, the process of deliberation is conceived exclusively as a means for each participating individual to reach their individual goals.[2] This is, for example, a goal of Harsanyi and Selten's (1988) monumental work on equilibrium selection, although as they themselves acknowledge, the goal is never reached.[3]

Towards its goal, the Bayesian procedure is to define a decision problem very abstractly, by its essentials only: preference structures, rules of play, and knowledge that these things are themselves common knowledge to all participating, who (incidentally) are assumed to be consummately rational beings, in possession of common knowledge of this mutual consummate rationality, as well as unbounded deductive powers. The participants are then assumed to pass from knowledge of the essential about their situation, by deliberately putting to one side (if necessary) such things as cultural norms and all other common knowledge not specifically relevant to the decision problem, via pragmatic reasoning, to a solution that compels action.[4] It shall be my thesis that, while the Bayesian approach expects the deliberator to be a consummate logician, as a consequence of being consummately rational, and in that regard exaggerates the human endowment, it at the same time understates her other resources.

2. Nash Equilibrium

The Bayesian approach to coordination (among other things) forms one stream of the so-called *Nash Program,* which commends solutions to dilemmas in the form of so-called *Nash equilibria.*[5] To define a Nash equilibrium we shall need to introduce the notion of best reply. A certain option for a given player is a *best reply* to other players' choices, if it best advances that player's aims (in Bayesian terms, maximizes his utility), in light of the other players' current choices. Then a Nash equilibrium is a combination of plays, one from each player, such that each one is a best reply on its player's part, to the other players' choices. A Nash equilibrium point is *self-enforcing* in the following sense: if each player believes that the others are doing their parts in the equilibrium stratagem, then each can be certain he can do no better for his aims by unilaterally deviating from equilibrium, and so will follow through with it, without there being a penalty imposed for deviation. For each player's part in the equilibrium is a best reply for each to what the others are doing. Now an option which is a best reply to other players' choices, *no matter what the others choose*, is called *dominant*. A dominant strategy is an unconditional best reply.

The sticking point is that there is more than one Nash equilibrium in typical games, none of them supportable purely on dominance considerations. For example,[6] suppose you and I are each given only one opportunity to name either Heads or Tails without communicating; we win a prize so long as our selections match. The outcome of the equilibrium point Heads/Heads is just as

agreeable to us as the outcome of the equilibrium Tails/Tails. The trick is to reach one equilibrium point, rather than the other, without benefit of communication.[7] The naive Bayesian approach, which simply directs its adherents to maximize individual utility, without giving further instructions in the matter, does not provide a mechanism for deliberating among competing equilibria, so does not prevent miscoordinations.

There is now a very large literature attempting to refine the Nash equilibrium concept, so that the refinement will both (1) eliminate intuitively incorrect equilibria,[8] and (2) leave only one equilibrium left standing as an option. For it was once thought that, if there is one unique equilibrium point, every theory, no matter its foundations, can embrace it as the solution to the game, since it has no competitors.[9] This is a quite problematic proposal, because it simply and unceremoniously *postulates* that all coordination problems possess a solution, however we like to think of what it is to be a solution. If we are Bayesians, this has the effect that every coordination problem possesses a solution that can be supported on individual utility-maximizing grounds. But

there are at the same time reasons for embracing the contrary of the existence postulate.[10] Michael Bacharach, who is perhaps the only person to axiomatize what may plausibly be called "the theory of noncooperative games" as practiced in the 1980's, has proved that a Bayesian theory *without* principles of deliberation beyond "Maximize!" does not have solutions to games with *unique* Nash equilibria.[11] The ground of this proof is quite simple, and is captured in an argument, put forward independently by Margaret Gilbert, against use of a salient property of a strategy combination (a property which allows that combination to draw prominent attention to itself as unique, and thereby to stand out as special) as potentially identifying that combination as a unique solution. The argument, in effect, is that even in the presence of a salient strategy combination, there is still nonetheless no reason for one player to believe that his partners or opponents will or are likely to perform *their* parts in that combination.[12] So if there is no reason for me to think you will choose your part in that prominent combination, I have no compelling reason, not even on grounds of maximizing utility, to choose my part in it. And so prominent considerations are not rationally compelling in the way they need to be in order to qualify as solutions.

One promising new approach is the idea that dilemmas ought be individuated also in terms of initial beliefs, that these too enter into the essentials of a decision situation. Brian Skyrms has perhaps the most prominent proposal, as well as the best worked out, along these lines, but Cristina Bicchieri has another.[13] The idea is that, in order to achieve a specific enough definition of a dilemma to reach a compelling solution, we must model also the knowledge of players, as to how they assess the *other* players are initially likely to play, and also to update systematically their opinions of each others' opinions, iterating the updating process until there is convergence.

The results of this approach are in fact interesting, but, as I now suggest, they are also of limited use. For the only factors it will allow as relevantly discriminating between one dilemma and another, are ones which impact players' initial assessment of each other's likelihood of pursuing the options open to.[14] The approach, in other words, requires that an initial assessment of the opponent or partner's play is ascribable to each player. These assessments are called *prior probabilities*, or simply *priors* (as they are known to friends and enemies alike). Allowing priors to individuate between games may be in keeping with the Bayesian program, but it runs into the following, fundamental, difficulty. If players fail, for whatever reasons, to form priors, or simply

fail to form them in a way which is not commonly known to be common knowledge, then such factors as would otherwise seem relevant cannot be used to discriminate one dilemma from another. And so the Skyrms proposal will have an application only in those cases in which factors that otherwise seem relevant to individuating one dilemma from another, *also* show up as making an impact on priors. So the proposal will succeed only if players have at least comparative opinion of each proposition to the effect that the opponent will do such-and-such, that these opinions are common knowledge, and that it is also common knowledge that players update this opinion systematically, according to some definite (and possibly also defensible) rule. But this is a tall order, even when a salient solution is known to be salient. For example in our Heads-Tails dilemma, each player might know that heads has a psychological first place for the normal citizen, without having any reason to think that *the opponent* is more likely to choose heads (Gilbert's point again).

Now Bayesians generally propose that we should simply manufacture priors, so long as they're consistent, if on introspection we don't find them on the shelf (as it were). One wishes to object here, for rather delicate reasons. For it is not that one is *against* the manufacture of priors, as such. How could one be? But rather on grounds that issuing an imperative to the effect that such priors must be manufactured is out of place, as an edict of prudence.[15] On at least two counts. First, if I do not estimate the chances that you will choose heads, how is it the business of a theory of *choice* to insist that I do? For how can this insistence be justified in terms of how my interests will be served? And if I can produce priors at will, why should I not produce them so that their magnitude is to my advantage, rather than apportioned according to the evidence? (The point is simply that epistemological imperatives cannot simply be a subcategory of practical imperative. A good illustration of this fact is the Predictor problem: it is in my interests to have a belief in, for example, backwards causation — if I can form such a belief far enough in advance that it might serve as grounds for the Predictor leaving a large monetary sum in the opaque box; it would be a belief of convenience, as we might then call it, but surely not one that would qualify as knowledge.) For isn't it the business of a theory of choice to tell me how to advance my interests, rather than my knowledge *as such*?

The second ground for objecting to the priors requirement, is simply that holding *any* prior whatever is liable to introduce logical inconsistencies into my corpus of beliefs, in the presence of a

common knowledge assumption. We might bring this point out as follows, *vis-a-vis* the Heads-Tails dilemma. Suppose I believe you are more likely to choose Heads than Tails, possibly on grounds that Heads is psychologically more prominent for you, and also that this is common knowledge between us. This belief, in the presence of a common knowledge assumption, presupposes that I have considered the impact of my own mind on the likelihood of your playing Heads. But this, in turn, can be true only if I know what I will do, or have some estimate of it, and this cannot be true unless I have already completed deliberations, or at least begun them, which has got to be false if I am forming a prior only.[16] Skyrms defenders are certain to object that I have confused priors with posteriors — thus confusing the *process* of deliberation with its end products — and that anyway inconsistencies tend to disappear as rational deliberation advances in real time. But my point is simply to ask why we should begin deliberations using priors in the first place, if they are the sorts of things that requires laundering.

The Bayesian might reply that the laundering of belief is feature of their prescriptive enterprise: they propose to explain what actually happens by drawing attention to what individuals *ought* to do in the interests of self-interest. For under the new scheme of explaining that Darwinian natural selection prescribes, the distinction between justifying a behavior and merely describing it, is either blurred or disappears. Now I agree that we can no longer explain social behavior without speaking about what (objectively) promotes the interests of individuals, since what we aspire to explain is not individuals' *actual* rationales for what they do, but rather what contributes to the *success* of the behavior in question. It does not follow, however, that we will be allowed to prescribe X to an individual, when something lesser, Y, will do equally well. I think something less than full-fledged belief will do, and therefore that Bayesian tolls are extortionate.

Now the Bayesians might simply insist that I always *do* have a degree of belief in the truth of something, whether I like it or not, if only I will reflect on the matter: I have only to Ramsify (think about how I would decide between certain gambles) and these exquisite items will materialize. So I'd best put them to use. But again: why is it the business of a theory of choice to insist that I bring these priors forward (assuming all they are on the shelf, and just require dusting off), and subsequently modify them according to a prescribed rule? Why are they relevant for individuating dilemmas when hair color, as such, or a salient property, as such, is not? Surely a theory of *choice* must have a prescription for me, even if I do not or cannot form a full comple

ment of priors, and should direct me to put aside all such beliefs as would foreseeably lead to an undesirable end.[17]

In summation, the demand for priors makes a virtue out of a necessity, which is in any case simply an artifact of the Kantian foundations of Bayesian theory, as I will now endeavor to explain.

3. Dependent Coordinations: An Anti-Kantian Proposal

The regularities in human social behavior rests on many things, among them the demands of rationality, or *prudence* as I prefer to call it. What are those demands? To my mind, decision theory as we know it today, because it has been shaped by Bayesian dogma, has made serious mistakes in articulating the nature of prudence. Primarily by embracing a Kantian view of agency.

Kant, as is well known, put forward the proposal that autonomy is a property of the rational will. Others, in the same spirit, have argued that it is self-contradictory to say of a nonautonomous will that it heeds reasons, and so is rational, since a nonatonomous will does not direct itself but *takes* direction from another agency (someone or something in authority over it), and so has no use for reasons.[18] Now it is this idea, that a rational individual is not one who is moved externally, on which rests the Bayesian presumption that an agent requires *positive reasons* in favor of one course of action over against another, and so requires a *deduction* of the solution to a dilemma. And this is the presumption I wish to challenge.

When two friends lose each other in a crowd, how do they proceed to find each other again? As we may say, it depends. For the answer to their problem cannot be reached *a priori,* but depends on the nature of their venture up to the point of getting lost. Each might find herself heading towards the point of initial meeting, if that is sufficiently close by. Or towards the destination point of their journey, if that is closer. Or some intermediate point, if it exerts more pressure on each of their sensibilities than either of the first two options. But if I lose my toddler, or my dog, on the *other* hand, I do not return to a departure or destination point, nor do I direct my path towards a prominent intermediate point, but instead seek out wandering places likely to attract the attention of a toddler or dog, close to the point of being separated.

I will draw two morals from these contrasting examples. The first is this: whether they succeed in coordinating or not, the friends choose as they do partly — but *only* partly — because

the option each chooses enjoys a psychologically more prominent position on their spectrum of perceived options. But it is the *rest* of the story as to why they choose as they do which is most interesting, philosophically speaking. I propose to tell it as follows: under circumstances when coordination with others is required, and there is a prominent, individuals who perceive the prominent strategy as prominent typically find themselves in possession of an *impulse* to act according to the prominent strategy. This impulse is a species of compulsion rather than a species of reason, which nevertheless admits of being either overridden or bypassed under appropriate conditions, as well as it admits of being complied with. In other words, I propose to understand certain features of circumstance as exerting an external power on agents, which give rise to conditional impulses to act. The properties of circumstance can give me, not a *reason* for acting, but instead an *impulse* towards a certain action or stratagem — an impulse whose force can be measured against that of reasons, as well also as against that of other impulses.[19] Thus when separated from *you*, I am moved to choose a prominent point *not* because I have reason to believe you will choose it, or a reason to believe you will think I will choose, or anything of the kind — although after the fact I am entitled to say that I believed you would do likewise because you had the same perspective on our situation. I choose my course of action simply because the point of reference exerts a certain pressure *directly* on my will, which I can elect to resist. I receive from my circumstances an impulse to act — an impulse which deserves compliance because I have no reasons to override it.[20]

The second moral I wish to draw from my examples concerns the differences when it comes to coordinating with a toddler or pet. I give in to the impulse to seek a peer at a place which is prominent for me, on grounds that I have no reason to believe that if I took my peer's perspective, I would be choosing differently. (But once again, these grounds are not reasons, if I do not articulate them to myself as such; nonetheless, they are the grounds of my action — the fonts or sources of my action.) But the story is quite different when it comes to being separated from my toddler. I must bypass the impulse I receive, as *myself,* in that case. And I must take his perspective. I neither model him as a maximizer of utility, just like myself, nor as a duplicator of my reasoning processes — although he is surely both, at least on many occasions. Instead I try to imagine the situation from his point of view, to take his perspective on the situation, confident that he is positively *not* perspective-taking himself. Perspective-taking is not the same thing as

modeling the other party after yourself-as-a-rational-being, although in the case of the two friends, the two enterprises are indistinguishable in their results.

I am therefore recommending perspective-taking, in the name of prudence, not (as the Bayesians counsel) the modeling of others after self-as-rational. Perspective-taking is what is called for when I seek to coordinate with my toddler; indeed I would and should be faulted for modeling him after myself. So why suppose it should not be required also when I seek to coordinate with my friend? And by advocating perspective-taking I *also* explain the significance of salience considerations for coordination, whilst agreeing that they cannot bear the entire burden of explaining why people choose in accordance with them. On my account, two friends who aspire to coordinate for the sake of being reunited, should not choose *purely* on grounds of prominence, or even on the assumption that the other (like self) sees a certain option prominently; I agree with Gilbert, that prominence is not enough *as a reason*. But it is enough to generate an impulse. Hence I am recommending the prominent solution precisely in those instances in which the prominent solution gives rise to the same impulse to act in all would-be coordinators, an impulse that is not overridden but instead reinforced through perspective-taking. Thus the grounds on which I say coordination is achieved are precisely those on which it is permitted by prudence: coordination rests, not purely on reasons, but on a combination of reasons and impulses.

Thomas Schelling in his groundbreaking work in strategy theory was, I believe, groping towards this anti-Kantian view of agency, but was nevertheless reluctant to make a clean break with the requirement of pure reasons. He wrote, for example, that certain signaling properties are the sorts of things that "seem to 'point toward' [one equilibrium point over against another]. They provide either a reason or an excuse for believing or pretending that [one equilibrium point is better than another]; since we need an excuse, if not a reason, for pretending, if not believing, that one of the equilibrium pairs is better, or more distinguished, or more prominent, or more eligible, than the other".[21]

Now the Bayesian might say that perspective-taking, if required at all by prudence, is required only instrumentally, for the sake of helping me achieve a *prediction* of my son's behavior, or my dog's, and so assists me establishing an opinion that will serve as a prior. This idea, to my mind, overintellectualizes the process I undergo when I coordinate with my son, so overintellectualizes what prudence requires. For I do not take my son's perspective in order to

predict his behavior. I do it for less intellectual reasons. I take his perspective in order to put myself in a position to be *directly* influenced by those same psychological factors as give rise to *his* impulses. It should come as no surprise — if indeed it is true — that perspective-taking is a good means to predicting his behavior. But asked to make a likelihood assessment of my son being found where I am looking — when in the grip of that electrifying fear for his safety, which compels me to take his perspective — I should certainly give the Bayesian no satisfaction. Can I be faulted, on grounds of prudence, for having made no such a likelihood assessment? If the answer is — as I believe it is — a negative one, then the Bayesian proposal requires too much as a theory of prudence: it makes a virtue out of a necessity that only a Kantian would accept as a necessity. And by the same token its model of the rational agent is overly intellectual.

The moral of my examples is therefore a caution against embracing Kantian autonomy unconditionally, as a precondition of rational decision making. It directs against using individual, utility-maximizing rationality as the sole grounds for commending a solution to players in a dilemma, and it rejects also the Nash equilibrium as a necessary property of a solution, much less a sufficient one. Thereby it opens the door to the possibility for utilizing such things as the Pareto Principle, which recommends replacing one strategy combination, however arrived at, by another which is preferred by some party but not dispreferred by any party to the problem. (Harsanyi and Selten call upon that principle, in a restricted form, in their account of equilibrium selection.) I have not — and will not — argue for the legitimacy of other principles at this time, as it would stray too far from the main point. Simply, I suggest that once the door to impulses is open, such things as the Pareto Principle might walk in, bidden or not, if only on its credentials as a prominent idea.

But a Pareto Principle is legitimate only if the enterprise at hand is viewed by those viewing it at least partly as collaborative, and not strictly as competitive. For, as Harsanyi and Selten are quite right to point out, it is a principle of what may be termed *collective rationality*. So if it is ever not imprudent to give in to an impulse to apply a Pareto Principle, then surely not all prudent decision-making can be viewed, even from a normative perspective, as the result of individual efforts summed "linearly" (as it were). Instead, coordination might be the result of collaborative effort, thought of not simply as resulting from communication, but resulting also from each member of the effort taking the perspective of everyone else — or, better, the result of everyone

taking a *group perspective*. By taking a group perspective, an individual is not acting autonomously, for she is taking guidance also from the aims of other, whether she embraces them herself or not. This is an Anti-Kantian proposal.[22]

I am thus suggesting that we need a theory of decision which individuates dilemmas also according to the relevant impulses at play, however they happen to arise, and with their proper disposition — whether they ought to be complied with or resisted. It will be a decision theory that does not place such a great premium on individual rationality, but might possibly also save an important place for group rationality and common aims to play, *as* common, in giving rise to legitimate impulses to act upon.

Thus I give support to John Searle's intuition, to the effect that many instances of goal-directed behavior, among the likes of us, cannot be thought of as a summation of individual aims at individual goals — that, in other words, the enterprise of aiming collectively is irreducible to configurations of aiming individually, with knowledge that others are so aiming as well. But whereas Searle rests his case for the irreducibility of collective aiming on the strength of an intuition, to the effect that we-intentions are irreducible to configurations of I-intentions,[23] I am suggesting we ought to embrace the irreducibility thesis because the *metaphysics* of atomic-level explanations, together with the proposal that the atoms are independently moved, is inadequate for handing such questions as what two friends, separated in a crowd, ought to do in order to be reunited, or what a parent ought to do for the sake of reuniting with a small child.

Searle's insight, with which I am in complete agreement, is that we-intentions can be more fundamental, metaphysically speaking, so that they give rise to I-intentions, and are not merely the sums of I-intentions, conceived of as primary and not derivative. In other words, that we-intentions can emerge from the primordial psychological ooze as we-intentions proper, standing on their own two feet as it were. And that I-intentions can be *dependent* upon we-intentions in an important metaphysical sense, as they flow from them. The idea being that, when two of us favor a coordination, and each knows of the other's favorable view of coordination, and each approves of the other's favorable view, and each knows of the other's favorable view of the favorable view, and so on — then each of us is in the best possible position for forming, on impulse, a we- intention, as such — an intention to act in concert with the other rather than to act unilaterally, to act as part of a larger action-taking body;[24] only subsequently and

derivatively does this intention proceed to give rise to I-intentions to do one's part in the collaborative enterprise. From this we-intention can flow a reason for acting in one particular way rather than another (for example, of choosing an equilibrium point which awards each of us more than any of the other equilibrium points), or if conditions favor it, an impulse towards a certain action. Provided there is no countervailing reason (or impulse), the I-intention in that instance will carry the day. And this proposal, on my view, is the only one that can account for the commonplace coordinations that a Bayesian account cannot explain.

What a we-intention does, subsequent to its formation, is to *correlate* the behaviors of collaborators, by giving rise to coordinated impulses for action or reasons to act.[25] This is how it functions to serve the interests of coordination. Thus, metaphysically speaking, what a we-intention does is reduce the number of degrees of freedom in, for example a two-by-two decision problem, from two degrees to one.

Ordinarily, when each of N persons has a decision to make, which impacts also what each of the other N-1 persons will receive as a result of these N decisions, we suppose that there are N *degrees of freedom* in the situation, one for each decision maker. But if the parties to such a decision problem seek to coordinate their behaviors perfectly, they will seek to choose *as if* they were a single decision-making body, so that the number of degrees of freedom is reduced from N to one (and thereby also the potential for miscoordination). This, at least, is what happens in the ideal cases, when the attempt to reduce degrees of freedom actually succeeds. When a group of N voters each has to cast a vote for one of two candidates, and there are two political parties each sponsoring a candidate, and there is widespread party loyalty, then efforts will be expended to reduce the number of degrees of freedom from N to 2. These efforts might not succeed entirely.

4. The Work of Reducing Degrees of Freedom

Reducing degrees of freedom is a piece of social engineering. As such it typically takes hard work on the parts of those who wish it done. The attempt can go wrong in many ways, many more ways in fact than it can go right. But it can also go right in a number of ways, as I wish now to illustrate.

When two separated friends seek to coordinate their actions for the sake of being reunited, they seek to act collaboratively. When they do, they reduce the number of degrees of freedom in their decision problem from 2 to 1, by *sharing* (as I will call it) the remaining degree of freedom;

they seek to act as if the two of them were each one half of a single decision-making unit. They seek for their choice to be *conditioned* by the choices others make; they seek to make their selection *dependently,* or nonautonously, in ways that Kant could never approve. By the same token, when I seek to be reunited with my lost toddler, who himself seeks to be found, we too seek to act collaboratively. And when we succeed in our effort, we reduce the number of degrees of freedom in our decision problem from 2 to 1. However, in our case we do it *not* by sharing the remaining degree of freedom equally, as friends equal to each other in strategic maturity do. Instead I, as the parent, relinquish any claim to a portion of the degree of freedom we aspire to share, to my child, who inadvertently, by default, claims it all for himself. For he cannot act as if one half of a single decision-making unit, due to his extreme youth. His claim to the whole of the single degree of freedom, though not unchallenged by my claim to it, nevertheless overrides mine. His deficiencies give him first title to it. This too is an idea that has been brilliantly discussed by Schelling, who tells us that formidable capability does not always translate into advantage, and weakness can oftentimes be better utilized to advantage than strength. For example, a nation state unable to stabilize its economy is in a position to achieve economic assistance it might otherwise be refused, and a weak intelligence can provide protection against threats.

5. Reasons and Causes

On the subject of the relations between reasons and actions, one recent view stands out among many: Donald Davidson's. According to Davidson's view, reasons and actions interact, by way of the first serving as causes of the second, and therefore as explanations, within a metaphysical framework composed of events and causal relations between them. I am proposing something quite different. I do not speak of events or causal relations for one very important reason: causal relations require there being events which stand in them, while events and their sums are not robust enough metaphysical entities to bear the weight of a science that handles the relations among reasons, decisions and actions.

On my account, reasons and actions interact only indirectly. Reasons are gatekeepers across the channel along which impulses pass, either checked or unchecked. I shall not say what impulses are, as I do not have to but can leave it up to others. Suffice it to say that impulses are not as such the things for which agents must take responsibility, as they are not under the direct

control of agents. Reasons, on the other hand, are the sorts of things for which we can hold agents accountable. They function either to obstruct the passage of an impulse, or let it pass unmolested, subsequently to contribute to action. Sometimes, though I dare say rarely, a reason can give rise to an impulse. As much as anything else, the job of maturation is to learn to harness impulses, of which there is no short supply, for the sake of our aims.

6. Metaphysics Against the Bayesian

The Bayesian agent is a consummate intellectual. His actions are based on reasons that have a foundation entirely in his corpus of beliefs. Agency, on the Bayesian account, consists of principles of choice applied to a corpus of belief; thus the whole business can be transacted propositionally, abstracted from the moral flesh and its waywardness. It is this intellectual approach, rather than the form of principles selected for it, which I am suggesting is mistaken. We cannot expect a purely intellectual system to perform as humans do, on grounds that intellectual systems always act *independently*, while flesh and blood can form nonintellectual alliances as well as intelectual ones — the flesh is subject to bonds which allow separate agencies to coalesce as their aims coincide. (This feature of my proposal therefore serves the foundational needs of classical economics, which is concerned with economic classes and their responsibilities for disposing of economic surplus, far better than the Bayesian foundations suggested by John Roemer.[26])

The ability to view matters from the point of view of someone else is not a purely intellectual enterprise, although it has its intellectual elements. It requires psychological capacities. No human knows what it's like to be a bat, and this reality is not due to purely intellectual deficits on our part: we simply do not have the right equipment to exercise, well or badly. For the knowledge in question is not purely intellectual. To know what it's like to be a bat, one must be in a position to *experience* the type of impulses a bat experiences, and this is only a small part of grasping the *agency* structure of a bat. The Bayesian proposal, to model others, as well as oneself, through propositions, therefore misrepresents the metaphysics of agency, simply by taking it off stage. I propose that the metaphysics of agency is central to understanding the successes of coordination, and so is central to understanding the nature of agency.

7. The Metaphysics of Agency

My proposal so far seems afflicted with contradiction. There appears to be a confusion of freedom with nonfreedom. When I say that parent and toddler share a degree of freedom, I give it to toddler not parent, on grounds that the toddler has a certain handicap, due to tender age, and therefore claims full title to the degree of freedom, while the parent by contrast is in possession of a capability the toddler does not have. And this seems to get things rather backwards. To answer this criticism, I shall have to discuss the metaphysics of freedom in my sense of the term — which is a sense borrowed partly from physics.

Freedom, in my sense, is not simply the capacity for putting some plan into action. If it were, there could be no reductions at all in degrees of freedom among full-fledged agents. Freedom in my sense, and as the physicists use the term, is the absence of constraint. Formally:

Def: X's decision is a *degree of freedom in a decision system S*, if and only if X's decision is among those factors that shape the state of system S.

The notion of shaping in this definition is used primitively, and is governed by the following axiom:

Ax: If a factor X shapes the state of a system S, or if it shapes another factor of system S, then it is false that X is given shape to by any other factor.

It is clear that, according to my account, there is no such thing as a degree of freedom, in the social world, absolutely: degrees of freedom are what they are only in relation to a system. But it is clear also that we do not require any entity such as a common mind, or a common body of knowledge, to define the notion of degree of freedom.

Now, lack of discipline (which is the toddler's handicap in our example, come by honestly through simple youth) gives one a claim to a degree of freedom. Whereas the ability to comply with an imperative, or to advance an aim, is no credential whatever, and is instead a counter-credential. For such an ability is precisely what I have in mind by constraint on action. It does not disqualify something's being my action, that I brought it about through following an imperative or seeking a particular aim, nor does it count against its being under my control, but it does count against my title to a full degree of freedom. When we count the number of degrees of freedom in a social system, we are not looking at the capacities for action on the parts of social players. We are looking, instead, at the number of *uncontrolled* or *unconstrained dimensions* of

action which can swing the social system towards one set of behaviors rather than another. Contrariwise, the degree of control or constraint in a social system is directly proportional to the level of social functioning within that system, which depends on the capacities for social functioning being exercised.

Reduction in degrees of freedom occurs among individuals with the capacity to see things from the point of view of others within the group. (But I do not claim that such a capacity is either necessary or sufficient for reduction in degrees of freedom.) And now we can see why the ability to see things from the point of view of others can help in reducing degrees of freedom: it helps with reducing the number of uncontrolled factors in the social system. When I look at things from the point of view of my toddler, I short-circuit the uncontrolled factors which arise from my own impulses, and allow his impulses to propel the social "system" towards the (in this case common) goal. Similarly when two political parties are vying for votes, the ability to see things from the point of view of others, either those in one's party or those in the other, enables one to disseminate the sort of information that will create the sorts of voting impulses one wishes created.

8. A Standing Objection

> Only individuals choose and act. Collectivities, as such, neither choose nor act and analysis that proceeds as if they do is not within the accepted scientific canon.[27]

It is no small minority that will press Buchanan's criticism that collectivities, as such, do not act, on grounds that collectivities, as such, have no preferences or wills. I reply that my account does not require that collectivities be possessed of preferences at all, or of collective wills as such.[28] My account requires only that the action of different wills be correlated, dependent. This correlation occurs as the result of such engineering as I described in the last section. And the idea that social engineering can reduce the number of uncontrolled social factors, which determine the dimensions of action in a social situation, is not at all problematic. In fact, the action of two wills can be correlated even if each randomizes, say, on the results of a coin toss, so long as it is the *same* coin which determines the result.[29]

8. More Delicate Matters

> ...the idea of supra-individual units of agency implies deep revisions to the conventional theory [of choice].[30]

This is not really true. The notions of Pareto optimality and utility superiority have been around for some time. It is just that the metaphysical framework for calling upon them has been missing; what was available made invoking them seem illegitimate — though some very respectable and mainstream theorists, like Harsanyi and Selten, have not hesitated. And the metaphysical idea of impulses being under the control of reasons has been around since David Hume. It is just that Kant's ideal of the rational agent as autonomous had banished all of the main ingredients of my account into a mist of oblivion. The proposal I am making, which suggests that we view certain impulses as being given rise to by principles such as Pareto optimality, unites the cast-off ideas into something that cannot be banished any longer as outcast.

One potential problem for an account such as mine, which admits of agents forming a supra-individual decision-making body, is to treat these supra-individual formations under the normativity of the theory. Is it a requirement of prudence that agents conceive of themselves as parts of a larger unit? For example, Susan Hurley argues that an adequate theory of rationality should address the question of "what the unit of agency, among those possible, should be".[31] On my view, the theory of prudence does not tell us what the appropriate unit for a decision problem should be; rather, its edicts are conditional upon the sort of units that actually form. This is because my account does not counsel as to what sorts of impulses should arise — this would be nonsense. Rather, it counsels as to whether the impulses that arise ought to be heeded, and does not counsel that only those be allowed to carry the day, which can be justified on individual utility-maximizing grounds alone.

[1] The original classical definition defined them also in terms of the permissibility of communication; this has come to be seen as both unnecessary and counterproductive.

[2] J. Buchanan, "Constitutional Economics," in Eatwell, Milgate, and Newman, eds, *The New Palgrave: A Dictionary of Economics* (London: Macmillan, 1987) 585-88, is perhaps most eloquent on the reasons: "Only individuals choose and act. Collectivities, as such, neither choose nor act and analysis that proceeds as if they do is not within the accepted scientific canon."

[3] *A General Theory of Equilibrium Selection in Games* (Cambridge: MIT Press, 1988). In the Postscript they write (p. 356): "This means that our theory uses two independent, and ostensibly very different, criteria of rationality. One of them, risk dominance, is based on *individual* rationality: it is an extension of Bayesian rationality from one-person decisions to n-person games involving strategic interaction among n players, each of them guided by Bayesian rationality. . . In contrast, payoff dominance is based on *collective* rationality: it is based on the assumption that in the absence of special reasons to the contrary, rational players will choose an equilibrium point yielding all of them higher payoffs, rather than one yielding them lower payoffs. That is to say, it is based on the assumption that rational individuals will cooperate in pursuing their common interests if the conditions permit them to do so."

[4] I will not survey the variety of solution concepts here, as it will take us too far afield. Suffice it to say simply that all solution concepts so far proposed, with the possible exception of correlated equilibrium, have had the minimum property of being Nash equilibria. A very sophisticated treatment of this issue is Michael

Bacharach, "A Theory of Rational Decision in Games," *Erkenntnis* 27 (1987), 17-55.

[5] This understanding of the Bayesian project results from recent harmonizations of classical game theory, with the Nash equilibrium solution concept at its core, and the Bayesian imperative to maximize individual utility, by R. Aumann, "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica* 55 (1987) 1-18, and B. Skyrms, *The Dynamics of Rational Deliberation* (Cambridge, 1990).

[6] This, and many other similarly compelling examples, first made appearance in Thomas Schelling's pathbreaking work, *The Strategy of Conflict* (Harvard University Press, 1960).

[7] In an important sense, communication does not really help. For if I am allowed to tell you (but not show you) that I am selecting Heads, how exactly does this give you a reason to choose Heads yourself?

[8] I shall not address the problems raised under this head.

[9] The monumental Harsanyi and Selten (1988), *op. cit.*, for example, rests on this idea, as do many other refinements of Nash equilibrium.

[10] Robert Sugden refers to it as *the Principle of Rational Determinacy* in "Rational Bargaining," in *Foundations of Decision Theory*, M. Bacharach and S. Hurley, eds

(Cambridge: Blackwell, 1991).

[11] "A Theory of Rational Decision in Games," *Erkenntnis* 27 (1987) 17-55.

[12] "Rationality and Salience," *Philosophical Studies* 57(1989) 61-77.

[13] B. Skyrms, *The Dynamics of Rational Deliberation* (Cambridge, 1990); J. Harsanyi, "The Tracing Procedure: A Bayesian Approach to Defining a Solution for n-person Non-cooperative Games," *International Journal of Game Theory* 4, 61-94; C. Bicchieri, *Rationality and Coordination* (Cambridge, 1993).

[14] The same holds true also of Bicchieri's distinct proposal (1993, *op.cit.),* although she does not employ priors as such.

[15] Richard Jeffrey's scruples notwithstanding.

[16] This resurrects an objection to priors that Skyrms believes he has answered to tolerable satisfaction, to the effect that it is illegitimate to make assessments of one's own likelihoods, to which Skyrms answers we need not. My point in the text is that a likelihood assessment of others, in the presence of common knowledge, presupposes a likelihood assessment of self.

[17] We should quite prepared, in other words, to find that the directives of prudence conflict with the directives of epistemology.

[18] Many have disagreed with Kant on this matter. Feminists, for example, disagree among themselves over whether autonomy is in any respect desirable, and whether women in particular ought to achieve autonomy wherever this is possible. There are feminists who believe that the trouble with autonomy is that women do not enjoy enough of it in real life, others who take a dim view of autonomy on grounds that it is tainted with (unworthy) masculine ideals, and still others who believe that autonomy is all too often confused with (both unrealistic and undesirable) separation from other human beings. See Christine di Stefano, "Trouble with Autonomy: Some Feminist Considerations," in *Feminism,* Susan Moller Okin and Jane Mansbridge, eds. (Vermont: Edward Elgar Publishing, Ltd., 1994), 383-404.

[19] This solves Gilbert's (1989) problem (*op. cit.),* by rejecting the proposition that only a *reason* for action can rationally justify a piece of behavior on grounds of prudence, and embracing instead the proposition that compliance with certain impulses is permissible so long as no compelling reason is raised *against.*

[20] This idea has some affinities with Aumann's notion of *correlated equilibrium,* although it admits as solutions strategy combinations, such as for example dominated strategy combinations, which Aumann will go to some lengths to rule out.

[21] *The Strategy of Conflict, op. cit.,* 297-8.

[22] The Kantian proposal, on the other hand, involves an irresolvable tension (as Kant himself saw clearly) between persons as totally autonomous, and persons as dependent on biological bodies which are subject to all manner of external influences.

[23] Searle says that his confidence in this irreducibility rests on an intuition, to the effect that all analyses of collective aimings at common goals, in terms of individual aimings at individual goals, plus common knowledge of such aims, will ultimately fail. And his reason for thinking this is that each of the individual aims which makes a contribution to achieving the common goal, when there is a common goal, is *derivative* — dependent for its existence on the common goal; and not the other way around. "Collective Intentions and Actions," in *Intentions in Communication*, P. Cohen, J. Morgan and M. Pollack, eds. (Cambridge: MIT Press, 1990), 401-415; reiterated in *The Construction of Social Reality* (New York: Free Press, 1995), 23ff.

[24] Bicchieri, *op. cit.*, views this iteration of knowledge and approval as an idealization, which may not hold good when we move away from a single decision maker. I view it as a perfectly reasonable and nonidealizing, since each individual is undertaking it, at this stage, from the point of view of a *single* decision-making body. But I agree it is quite an idealization if the individuals undertaking it are Bayesian atoms.

[25] Here is where my view comes close to the idea offered by R. Aumann, *op. cit.*, although, as I say, he does not give it the wide application I shall do; nor does he use

the concept of correlated action to replace the need for priors.

[26] *Analytic Foundations of Marxian Economic Theory* (Cambridge University Press, 1981); "Methodological Individualism and Deductive Marxism," *Theory and Society* 11 (1982) 513-20.

[27] J. Buchanan, "Constitutional Economics," in Eatwell, Milgate, and Newman, eds, *The New Palgrave: A Dictionary of Economics* (London: Macmillan, 1987) 585-88.

[28] M. Gilbert, *On Social Facts* (London: Routledge, 1989) is prepared to give an account of collective desires and beliefs. I am not.

[29] An example due to R. Aumann, *op. cit.*

[30] M. Hollis and R. Sugden, "Rationality in Action," *Mind* 102 (1993) 1-35.

[31] *Natural Reasons* (Oxford, 1989), 145.