AGENCY EQUILIBRIUM[1]

JONATHAN NEWTON[a]

Agency may be exercised by different entities (e.g. individuals, firms, households). A given individual can form part of multiple agents (e.g. he may belong to a firm and a household). The set of agents that act in a given situation might not be common knowledge. We adapt the standard model of incomplete information to model such situations.

KEYWORDS: agency, equilibrium, individualism, collectivism.

1. INTRODUCTION

The doctrine of *methodological individualism* (see Heath, 2015) is that social phenomena should be explained with reference to the actions of individual agents and the *intentional states* (e.g. belief, desire, hope, intention) that motivate these actions. Early proponents of methodological individualism (Schumpeter, 1909; Weber, 1922) regarded individual people (henceforth, *individuals*) as the only entities that could hold intentional states and consequently held that social theory should be founded upon individual agency. However, in practice, social scientists often model collective entitites (households, firms) or intra-human entities (multiple selves) as agents.[1] This is consistent with methodological individualism if such agents can hold intentional states or, in the sense of positive methodology (Friedman, 1953; Keynes, 1890), can act as if they hold intentional states.[2,3]

[1]This work is part of the shared intentions agenda that examines collective agency and social structure. For more, go to http://sharedintentions.net. I thank Atsushi Kajii and Toru Suzuki for comments, as well as the Kyoto Institute of Economic Research, which hosted me while the paper was written.

[a]No permanent address.jcsnewton@gmail.com

[1]When it comes to explaining social phenomena, the success of even this relaxed approach is a matter of dispute. See Arrow (1994) for a sceptical perspective, Samuelson (2016) for an enthusiastic one.

[2]For discussions of jointly held intentional states, in particular the case where the intentional state is an intention, see Searle (1990); Tomasello (2014); Tuomela and Miller (1988).

[3]It is possible to ask under what conditions evolution will favour the ability of individuals to occasionally act as if they hold joint intentional states (Angus and Newton, 2015).

The above raises the prospect of an individual forming part of multiple agents with conflicting interests. Consider, for example, a firm with two partners, Alice and Bob, who sometimes make decisions together with the goal of maximizing the firm's profits, but who also sometimes make decisions as singletons with the goal of furthering their own interests. Such conflicts of interest define the Prisoner's Dilemma (see Kuhn, 2014) and the Tragedy of the Commons (Hardin, 1968; Lloyd, 1833). Furthermore, an individual might not know the agents to which other individuals belong. Colm may know that Alice and Bob are partners, but in some specific situation be unsure over whether they will make a decision together or make separate decisions. Colm is unsure about whether he faces one other agent or two other agents.

Here we extend the concept of Bayesian Nash Equilibrium (Harsanyi, 1967, 1968a,b) to deal with incomplete information about the identity of agents in a game. In *agency equilibrium*, a random state of the world determines not only payoffs, but also a set of agents whose incentive constraints must be satisfied. Each agent comprises either a single individual (individual agency) or a set of individuals (collective agency). If the set of agents equals the set of singletons at every state, then the solution reduces to Bayesian Nash Equilibrium.

The basic model is very similar to that of Bacharach (1999). However, the primary interest of the cited paper is in situations where the membership of a collective (a 'team') is unreliable. For example, Alice may reason and act according to the perspective of a firm she co-owns with Bob, even though Bob is unreliable and sometimes reasons from an individual perspective. In contrast, here we focus on the case in which individuals within a collective agent are reliable. That is, both Alice and Bob know whether they are making decisions together or as individuals and act accordingly. The subtlety arises from the uncertainty of others over the agency of Alice and Bob. Bacharach (1999) can thus be seen as an analysis of individuals reasoning in isolation but taking the perspective of a collective[4], whereas the current model considers the strategic implications of the uncertain possibility of explicit face-to-face collective

[4]For more on team reasoning, see Bacharach (2006); Sugden (1993, 2003) and for a discussion of the relationship between team reasoning and collectivity in intentional states, see Gold and Sugden (2007).

deliberation.

The implications of uncertainty over agency are explored across several examples. Firstly, in a model of Cournot oligopoly, the possibility that other firms might have formed a secret cartel leads to singleton firms increasing their production, so that even in the absence of cartels, total production is higher than it would be in the model with complete information. Varying the probability of a cartel forming causes the expected quantity produced in equilibrium to vary continuously between the complete information outcomes for different numbers of firms. In a second example, we see that in coordination games, the possibility that other individuals are solving the coordination problem directly through explicit agreement can eliminate inefficient equilibria, even at states at which there is no collective agency. In a third example, it is shown that Nash equilibria can be non-robust to even vanishingly small amounts of uncertainty over agency. That is, even the smallest possibility of collective agency may suffice to ensure that the agency equilibrium outcome is far from the Nash equilibrium outcome of the game under complete information. Finally, a variant of agency equilibrium, *Pareto agency equilibrium*, is defined that relies upon Pareto dominance arguments and explicitly eschews interpersonal comparability of payoffs. This is explored through an extended version of the battle of the sexes game.

The paper is organized as follows. Section 2 describes games with incomplete information over agency, then defines and discusses agency equilibrium. Section 3 applies the concept to several examples. Section 4 defines and discusses Pareto agency equilibrium.

2. MODEL

An incomplete information game $\mathscr{U}$ consists of (1) the set of individuals, $\mathcal{I} = \{1, \ldots, I\}$; (2) the individuals' action sets, $A_1, \ldots, A_I$; (3) a countable state space, Ω; (4) a probability measure on the state space, P; (5), for each individual i, a partition of the state space, $\mathcal{Q}_i$; and (6), for each individual i, a bounded state dependent payoff function, $u_i : A \times \Omega \to \mathbb{R}$, where $A = A_1 \times \ldots \times A_I$. Thus $\mathscr{U} = \{\mathcal{I}, \{A_i\}_{i \in \mathcal{I}}, \Omega, P, \{\mathcal{Q}_i\}_{i \in \mathcal{I}}, \{u_i\}_{i \in \mathcal{I}}\}$. We write $P(\omega)$ for the probability of the singleton event $\{\omega\}$ and $Q_i(\omega)$ for the (unique) element of $\mathcal{Q}_i$ containing ω. We restrict attention to games where every information set of every individual is possible, that

is $P[Q_i(\omega)] > 0$ for all $i \in \mathcal{I}$ and $\omega \in \Omega$. Hence, the conditional probability of state ω given information set $Q_i(\omega)$, written $P[\omega|Q_i(\omega)]$, is well-defined by the rule $P[\omega|Q_i(\omega)] = P(\omega)/P[Q_i(\omega)]$.

This description is a standard description of a game of incomplete information (see, e.g. Kajii and Morris, 1997). However, for our purposes, we require two further objects.

(7) An *agency correspondence* $\Pi : \Omega \rightrightarrows \mathcal{P}(\mathcal{I})$ such that $\Pi(\omega)$ is a partition of $\mathcal{I}$ into *agents*. As a set of agents $\Pi(\omega)$ is a partition of $\mathcal{I}$, we have that, at any given state ω, individual i is a member of exactly one agent. Further restrict Π so that if $i \in J \in \Pi(\omega)$ and $\omega' \in Q_i(\omega)$, then $J \in \Pi(\omega')$. That is, the agent to which individual i belongs is constant across states in information set $Q_i(\omega)$, $\omega \in \Omega$.

(8) For each agent $J \subseteq \mathcal{I}$, a bounded state dependent payoff function, $v_J : A \times \Omega \to \mathbb{R}$, such that $v_{\{i\}} = u_i$ for all $i \in \mathcal{I}$. This specifies the payoffs according to which agents make decisions. It is restricted so that when an agent is a single individual, the agent's payoff function is identical to that of the individual concerned. In some instances it will be natural to define v_J in terms of u_i, for example when agents have utilitarian payoffs $v_J(a, \omega) = \sum_{i \in J} \lambda_i(\omega) u_i(a, \omega)$, where $\lambda_i(\omega)$ are strictly positive, possibly state-dependent welfare weights.

For $J \subseteq \mathcal{I}$, write $A_J = \prod_{i \in J} A_i$. Let $\mathcal{Q}_J$ be the coarsest common refinement of $\mathcal{Q}_i$, $i \in J$. That is, $\mathcal{Q}_J$ is a pooling of the information that individuals in J have about the state of the world. A *strategy* for agent $J \subseteq \mathcal{I}$ is a $\mathcal{Q}_J$-measurable function $\sigma_J : \{\Omega : J \in \Pi(\omega)\} \to \Delta A_J$. We write Σ_J for the set of such strategies. A *strategy profile* is a function $\sigma : \Omega \to \Delta A$ such that $\sigma(\omega) = (\sigma_J(\omega))_{J \in \Pi(\omega)}$, where σ_J is a strategy for agent J. Let Σ denote the set of all strategy profiles. Given ω such that $J \in \Pi(\omega)$, we write $\sigma_{-J}(\omega)$ for $(\sigma_K(\omega))_{K \in \Pi(\omega) \setminus \{J\}}$. When no confusion arises, we extend the domain of each v_J to mixed strategies and thus write $v_J(\sigma(\omega), \omega)$ for $\sum_{a \in A} v_J(a, \omega)\sigma(a|\omega)$.

Definition 1 Given game $\mathcal{U}$, agents' payoffs v, strategy profile $\sigma \in \Sigma$, state of the world $\omega \in \Omega$, then $a_J \in A_J$ is a *profitable deviation* for $J \subseteq \mathcal{I}$

if

$$\sum_{\omega' \in Q_J(\omega)} v_J(\{a_J, \sigma_{-J}(\omega')\}, \omega')\, P[\omega' \mid Q_J(\omega)] > \sum_{\omega' \in Q_J(\omega)} v_J(\sigma(\omega'), \omega')\, P[\omega' \mid Q_J(\omega)].$$

Definition 2 A strategy profile σ is an *Agency Equilibrium* of $(\mathcal{U}, \Pi, v)$ if, for all $\omega \in \Omega$, no profitable deviation exists for any $J \in \Pi(\omega)$.

Remark 1 If σ is an Agency Equilibrium of $(\mathcal{U}, \Pi^{NE}, v)$, where $\Pi^{NE}(\omega) = \{\{1\}, \{2\}, \ldots, \{I\}\}$ for all $\omega \in \Omega$, then $(\sigma_{\{i\}})_{i \in \mathcal{I}}$ is a *Bayesian Nash Equilibrium* (Harsanyi, 1968a) of $\mathcal{U}$.

Remark 2 The following two assumptions of the model are parallel. (i) An individual can potentially belong to many agents whose incentives may differ, but is only part of one agent at any given state. (ii) An individual can potentially have many preferences which may contradict one another, but has only one set of preferences at any given state.

Remark 3 In agency equilibrium, an agent $J \subseteq \mathcal{I}$ is *instrumentally rational* (Weber, 1922) in that the strategy of J is chosen to achieve the "goals, desires, and ends" (Nozick, 1994) of J as represented by v_J. As Nozick notes, "About the goals themselves, an instrumental conception has little to say." In our examples, we let v_J depend in a positive way on u_i, $i \in J$, but neither this, nor any other dependence on the individual payoffs is required by the model.

Remark 4 The individuals within agent J can be regarded as the instruments of J with respect to the instrumentally rational intent of J to maximize v_J. This highlights the independence of agency and payoffs. For example, if $v_J = \sum_{i \in J} u_i$, then the payoffs of J are entirely determined by the payoffs of its constituent individuals, yet the actions of the constituent individuals are entirely determined by J's instrumentally rational pursuit of these payoffs.

Remark 5 Given a state of the world, our model only allows an individual to be instrumentalized by a single agent. Hypothetically, if an

individual were to be simultaneously (i.e. at a given state) instrumentalized by more than one agent, then there would have to be no conflict of interest between these agents regarding the desired action of the individual concerned. From this perspective, it is without loss of generality to restrict individuals to be part of only a single agent at any given state. When this is not the case, for example in the Strong Equilibrium concept of Aumann (1959) in which every subset of $\mathcal{I}$ simultaneously exercises agency, equilibria may fail to exist due to conflicting incentives between agents. The same applies to Strong Equilibrium with incomplete information over payoffs (Ichiishi and Idzik, 1996) and to ideas of Coalition Proofness Bernheim et al., 1987; Moreno and Wooders, 1996).

REMARK 6 If mixed strategies are disallowed so that $\sigma_J(\omega)$ always puts all probability weight on a single action profile, then the model is a special case of the *unreliable team interaction* model of Bacharach (1999) under the restriction that if individual i reasons and acts from the perspective of collective J, then all individuals $j \in J$ reason and act from the perspective of J, and this is common knowledge amongst the individuals in J.

REMARK 7 An alternative approach to modeling multiple agency is to explicitly model adaptive dynamics in which a given individual may be part of different agents at different times (Feldman, 1974; Green, 1974; Newton, 2012a,b; Roth and Vande Vate, 1990). Such models give a complete description of behavior both in and out of equilibrium.

3. APPLICATIONS

3.1. *Cournot oligopoly*

Consider quantity competition between three firms[5], $\mathcal{I} = \{1, 2, 3\}$. Each firm $i \in \mathcal{I}$ chooses a production quantity $A_i = \mathbb{R}_{\geq 0}$. Individual firms' payoffs depend only on the quantities chosen by the firms and not

[5]Given the topic of the current paper, it is interesting to note that the original presentation of this model in Cournot (1838) does not mention firms. Instead the decision makers are the two owners of two springs: 'Maintenant, imaginons deux propriétaires et deux sources, dont les qualités sont identiques'.

on the state of the world, $u_i(a, \omega) = a_i(R - \sum_j a_j)$, $R > 0$. Agents' payoffs are utilitarian, $v_J(a, \omega) = \sum_{i \in J} u_i(a, \omega)$ for all $J \subseteq I$.

The state space is $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3\}$. At state ω_0, no firms form a cartel, $\Pi(\omega_0) = \{\{1\}, \{2\}, \{3\}\}$. At state ω_i, $i = 1, 2, 3$, firm i produces on his own and the other two firms form a cartel, $\Pi(\omega_i) = \{\{i\}, \{j, k\}\}$, $j, k \neq i$. Let each of the three pairs that can form a cartel do so with probability $p < 1/3$, so $P(\omega_0) = 1 - 3p$, $P(\omega_i) = p$ for $i = 1, 2, 3$.

The information structure is such that if firm i is not part of a cartel, then firm i does not know whether or not firms j and k form a cartel. Firms that are part of a cartel know this and hence know the exact state of the world. Therefore, $\mathcal{Q}_i = \{\{\omega_0, \omega_i\}, \{\omega_j\}, \{\omega_k\}\}$.

Consider an Agency Equilibrium which is symmetric in that both members of any cartel produce the same quantity. For $i = 1, 2, 3$, $i \notin J = \{j, k\}$, we obtain equilibrium strategies $\sigma_i(a_0 \,|\, \omega_0) = 1$, $\sigma_i(a_i \,|\, \omega_i) = 1$, $\sigma_J((a_j, a_k) \,|\, \omega_i) = 1$, where

$$a_0 = a_i = \frac{2 - 5p}{8 - 21p}, \qquad a_j = a_k = \frac{3 - 8p}{16 - 42p}.$$

At state ω_i, firms j and k form a cartel and reduce their production. At states ω_0, ω_i, firm i does not know whether j and k are in fact part of a cartel, but knowing that this is a possibility, increases production as a consequence. Note that at state ω_0, at which there is no cartel, every firm produces more than the Nash equilibrium quantity of $1/4$. As $p \to 0$, $P(\omega_0) \to 1$, cartels become a rare occurence and $a_0 \to 1/4$. Conversely, as $p \to 1/3$, we have that $P(\omega_0) \to 0$ and a cartel is almost a certainty. Production by both solo firms and cartels then approaches the Nash equilibrium quantity of the two firm model, $a_i \to 1/3$, $a_j + a_k \to 1/3$.

3.2. *Coordination game: Hi-Lo*

Let there be three individuals, $\mathcal{I} = \{1, 2, 3\}$ with action sets $A_i = \{H, L\}$. For each state of the world ω, let $u_i(a, \omega) = 2$ (respectively, 1) if $a_i = H$ (respectively, L) and $a_j = a_i$ for some $j \neq i$. Otherwise, $u_i(a, \omega) = 0$. That is, an individual obtains payoff from successfully coordinating on H or L with at least one other individual, with coordination on H giving a payoff of 2 and coordination on L giving a payoff of 1. Agents' payoffs are utilitarian with arbitrary state-dependent weightings,

$v_J(a,\omega) = \sum_{i\in J} \lambda_i(\omega)u_i(a,\omega)$ for all $J \subseteq I$, where $\lambda_i(\omega) > 0$ for every $i \in J$, $\omega \in \Omega$.

As in the previous example, the state space is $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3\}$, $\Pi(\omega_0) = \{\{1\},\{2\},\{3\}\}$, and for ω_i, $i = 1,2,3$, $\Pi(\omega_i) = \{\{i\},\{j,k\}\}$, $j,k \neq i$. Also as in the previous example, let $P(\omega_0) = 1 - 3p$, $P(\omega_i) = p$ for $i = 1,2,3$; $\mathcal{Q}_i = \{\{\omega_0,\omega_i\},\{\omega_j\},\{\omega_k\}\}$ for $i = 1,2,3$.

At state ω_i, $i = 1,2,3$, in agency equilibrium, agent $\{j,k\}$ will maximize $v_{\{j,k\}}$ by choosing $a_j = H$, $a_k = H$. It remains to determine equilibrium actions for individual i, $i = 1,2,3$ when i is a singleton agent, that is when the state is either ω_0 or ω_i.

For action L to be played in equilibrium, it must be the case that the expected payoff of player i at information set $\{\omega_0,\omega_i\}$ when he plays L is at least as high as his payoff when he plays H. His expected payoff from playing L is bounded above by the probability $(1-3p)/(1-2p)$, given by Bayes' rule, that the other two individuals do not form a collective agent. His expected payoff from playing H is bounded below by $2p/(1-2p)$, the probability that the other two individuals do form a collective agent multiplied by the payoff of 2 for successful coordination on H. This second quantity is greater than the first quantity if $p > 1/5$. Therefore, if $p > 1/5$, there is a unique agency equilibrium at which every individual plays H at every information set.

3.3. *Robustness of Nash equilibrium*

Here we illustrate the potential non-robustness of Nash Equilibrium to small amounts of uncertainty over agency. We combine three player matching pennies (specifically, Example 3.1 of Kajii and Morris, 1997) with a public goods problem. The cited example showed non-robustness of Nash Equilibrium to small amounts of incomplete information over payoffs. Here, there is no incomplete information over payoffs but there is incomplete information over agency.

Let $\mathcal{U}$ have three individuals, $\mathcal{I} = \{1,2,3\}$, with action sets $A_i = \{C,H,T,S\}$. Actions H and T are heads and tails in a matching pennies game in which individual i wishes to anti-coordinate with individual $\rho(i)$, where $\rho(3) = 2$, $\rho(2) = 1$, $\rho(1) = 3$. Actions C and S opt out of this coordination problem. In addition, action C also involves making a (non-individually rational) contribution to a public good. If we ignore the pub-

lic good component, C is regarded by other players as equivalent to H. Let $\Omega = \{\omega_i\}_{i\in\mathbb{N}_+}$ and $P(\omega_i) = (1 - \sqrt{1-\varepsilon})(\sqrt{1-\varepsilon})^i$. Partitions are

$$\begin{aligned}
\mathcal{Q}_1 &= \{\{\omega_1,\omega_2,\omega_3,\omega_4\},\{\omega_5,\omega_6,\omega_7\},\{\omega_8,\omega_9,\omega_{10}\},\ldots\},\\
\mathcal{Q}_2 &= \{\{\omega_1,\omega_2,\omega_3,\omega_4\},\{\omega_5\},\{\omega_6,\omega_7,\omega_8\},\{\omega_9,\omega_{10},\omega_{11}\},\ldots\},\\
\mathcal{Q}_3 &= \{\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\},\{\omega_7,\omega_8,\omega_9\},\{\omega_{10},\omega_{11},\omega_{12}\},\ldots\}.
\end{aligned}$$

Let $\mathfrak{C}(a_{-i})$ be the number of individuals, excluding individual i, who play C at action profile a. That is, $\mathfrak{C}(a_{-i}) = |\{j \in \mathcal{I} :\ j \neq i, a_j = C\}|$. Let

$$\begin{aligned}
u_i(\{C,a_{-i}\},\omega) &= 9\,\mathfrak{C}(a_{-i}),\\
u_i(\{S,a_{-i}\},\omega) &= 9\,\mathfrak{C}(a_{-i}) + 1\\
u_i(\{H,a_{-i}\},\omega) &= \begin{cases} 9\,\mathfrak{C}(a_{-i}) - 4 & \text{if } a_{\rho(i)} \in \{C,H\}\\ 9\,\mathfrak{C}(a_{-i}) + 4 & \text{if } a_{\rho(i)} = T\\ 9\,\mathfrak{C}(a_{-i}) & \text{if } a_{\rho(i)} = S \end{cases},\\
u_i(\{T,a_{-i}\},\omega) &= \begin{cases} 9\,\mathfrak{C}(a_{-i}) + 4 & \text{if } a_{\rho(i)} \in \{C,H\}\\ 9\,\mathfrak{C}(a_{-i}) - 4 & \text{if } a_{\rho(i)} = T\\ 9\,\mathfrak{C}(a_{-i}) & \text{if } a_{\rho(i)} = S \end{cases}.
\end{aligned}$$

Note that individual i always obtains a higher payoff from $\{S,a_{-i}\}$ than from $\{C,a_{-i}\}$ and that the individual best responses of individual i are T, T, H, S when $\rho(i)$ plays C, H, T, S respectively.

Consider an agency correspondence Π such that if $\omega \in \{\omega_1,\omega_2,\omega_3,\omega_4\}$, then $\Pi(\omega) = \{\{1,2\},\{3\}\}$. If $\omega \notin \{\omega_1,\omega_2,\omega_3,\omega_4\}$, then $\Pi(\omega) = \{\{1\},\{2\},\{3\}\}$. Recall that $v_{\{i\}} = u_i$ for $i = 1,2,3$, and let $v_J = \sum_{i\in J} u_i$ for $J \subseteq \mathcal{I}$.

At Agency Equilibrium of $(\mathcal{U},\Pi,v)$,

$$\text{If } \ \omega \in \{\omega_1,\omega_2,\omega_3,\omega_4\}, \ \text{ then } \ \sigma_{\{1,2\}}(\,(C,C)\,|\,\omega\,) = 1,$$

otherwise $a_{\{1,2\}} = (C,C)$ would be a profitable deviation for agent $\{1,2\} \in \Pi(\omega)$. Unique individual best responses then dictate that

$$\begin{aligned}
&\text{If } \ \omega \in \{\omega_1,\ldots,\omega_6\}, \ \text{ then } \ \sigma_3(\,T\,|\,\omega\,) = 1;\\
&\text{If } \ \omega \in \{\omega_5,\omega_6,\omega_7\}, \ \text{ then } \ \sigma_1(\,H\,|\,\omega\,) = 1;\\
&\text{If } \ \omega \in \{\omega_5,\omega_6,\omega_7,\omega_8\}, \ \text{ then } \ \sigma_2(\,T\,|\,\omega\,) = 1;\\
&\text{If } \ \omega \in \{\omega_7,\omega_8,\omega_9\}, \ \text{ then } \ \sigma_3(\,H\,|\,\omega\,) = 1;
\end{aligned}$$

and so on. Action S is never played, even for arbitrarily small values of ε.

In contrast, in Nash Equilibrium of $\mathcal{U}$, action C is never played as it is strictly dominated by S. It is left to the reader to check that the unique Nash equilibrium of $\mathcal{U}$ has S being played by every individual at every information set. Hence even small amounts of incomplete information over agency can dramatically change the implications of equilibrium analysis.

4. EQUILIBRIUM WITHOUT COLLECTIVE PAYOFFS

Agents' payoffs, v_J, $J \subseteq \mathcal{I}$, in our examples above involve interpersonal comparison of payoffs. From a revealed preference perspective, such interpersonal comparison is implicit in the choices made by a collective agent. However, from an ex-ante perspective, we may wish to define a variant of agency equilibrium that uses individual payoffs as the building blocks for aggregate behaviour, but does not make interpersonal comparisons. To this purpose, we now drop agents' payoffs from the model and instead let the choices of a collective agent J be guided by Pareto comparisons between vectors of individual payoffs, $\{u_i\}_{i\in J}$.

Definition 3 Given game $\mathcal{U}$, strategy profile $\sigma \in \Sigma$, state of the world $\omega \in \Omega$, then $\tilde{\sigma}_J(\omega) \in \Delta A_J$ is a *Pareto deviation* for $J \subseteq \mathcal{I}$ if, for all $i \in J$,

$$\sum_{\omega' \in Q_J(\omega)} u_i(\{\tilde{\sigma}_J(\omega), \sigma_{-J}(\omega')\}, \omega')\, P[\,\omega' \mid Q_J(\omega)] \geq \sum_{\omega' \in Q_J(\omega)} u_i(\,\sigma(\omega'), \omega')\, P[\,\omega' \mid Q_J(\omega)]$$

with strict inequality for some $i \in J$.

Definition 4 A strategy profile σ is an *Pareto Agency Equilibrium* of $(\mathcal{U}, \Pi)$ if, for all $\omega \in \Omega$, no Pareto deviation exists for any $J \in \Pi(\omega)$.

Note that, by Definition 3, Pareto agency equilibrium explicitly requires robustness to deviations which are mixtures of action subprofiles $a_J \in A_J$. With agency equilibrium (Definition 2), it was not necessary to include mixed deviations, as if no profitable pure deviation exists according to Definition 1, then no profitable mixed deviation exists. This is

	O	F
O	**3, 3, 4**	6, 0, 4
F	0, 6, 4	0, 0, 0

O

	O	F
O	0, 0, 0	0, 4, 6
F	4, 0, 6	**2, 2, 6**

F

FIGURE 1.— Battle of the sexes with 2-1 sex ratio.

not true for Pareto deviations, where for some agent $J \subseteq \mathcal{I}$, it may be that neither a_J nor a'_J are Pareto deviations, but some mixture of a_J and a'_J is a Pareto deviation. This is a potential criticism of mixed strategies in a multi-agency setting, as Pareto deviations for J may include action subprofiles a_J in their support that on their own would lead to a reduction in expected payoff for some $i \in J$. This implies that robustness to Pareto deviations is a strong robustness requirement. Equilibria that are robust to mixed deviations are, a fortiori, robust to deviations in pure strategies.

4.1. *Example: Battle of the sex ratio*

Figure 1 illustrates the battle of the sex ratio (Newton, 2012a), a three player version of the battle of the sexes. Two players, the row and column players, prefer opera (O) to football (F) and wish to coordinate with the matrix player. If they both coordinate with the matrix player, then they have an equal chance of accompanying the matrix player to the chosen event, and thus share the coordination payoff. The matrix player prefers football to opera and wishes to coordinate with at least one of the other players.

Let the set of individuals be $\mathcal{I} = \{1, 2, 3\}$ with action sets $A_i = \{O, F\}$. Let individuals $1, 2, 3 \in \mathcal{I}$ respectively correspond to the row, column and matrix players of the game in Figure 1. For each state of the world ω, let individual payoffs u_i be given by the entries of the payoff matrix. Consider a state space, $\Omega = \{\omega_0, \omega_1\}$, $\Pi(\omega_0) = \{\{1\}, \{2\}, \{3\}\}$, $\Pi(\omega_1) = \{\{1\}, \{2, 3\}\}$. Let $P(\omega_0) = 1 - p$, $P(\omega_1) = p$; $\mathcal{Q}_1 = \{\{\omega_0, \omega_1\}\}$, $\mathcal{Q}_2 = \mathcal{Q}_3 = \{\{\omega_0\}, \{\omega_1\}\}$.

There are three possible Pareto agency equilibria in pure strategies. For two of these equilibria to exist, the probability p that the collective agent $\{2, 3\}$ is formed (i.e. the state is ω_1) must be low enough that the actions chosen by the collective agent do not affect choices when the collective

agent does not form (i.e. at state ω_0). For $p \leq {}^3\!/_5$,

$$\sigma_1(O \,|\, \omega_0) = \sigma_1(O \,|\, \omega_1) = 1, \quad \sigma_2(O \,|\, \omega_0) = 1, \quad \sigma_3(O \,|\, \omega_0) = 1,$$
$$\sigma_{\{2,3\}}((F,F) \,|\, \omega_1) = 1,$$

is an equilibrium, and for $p \leq {}^2\!/_5$,

$$\sigma_1(F \,|\, \omega_0) = \sigma_1(F \,|\, \omega_1) = 1, \quad \sigma_2(F \,|\, \omega_0) = 1, \quad \sigma_3(F \,|\, \omega_0) = 1,$$
$$\sigma_{\{2,3\}}((O,O) \,|\, \omega_1) = 1,$$

is an equilibrium. The third equilibrium exists for all $0 < p < 1$. This is the equilibrium at which action F is always played by every agent,

$$\sigma_1(F \,|\, \omega_0) = \sigma_1(F \,|\, \omega_1) = 1, \quad \sigma_2(F \,|\, \omega_0) = 1, \quad \sigma_3(F \,|\, \omega_0) = 1,$$
$$\sigma_{\{2,3\}}((F,F) \,|\, \omega_1) = 1.$$

We shall see below that there is no equivalent equilibrium at which action O is always played.

Pareto agency equilibria in mixed strategies also exist. For example,

$$\sigma_1(F \,|\, \omega_0) = \sigma_1(F \,|\, \omega_1) = 1, \quad \sigma_2(F \,|\, \omega_0) = 1, \quad \sigma_3(F \,|\, \omega_0) = 1,$$
$$\sigma_{\{2,3\}}((F,F) \,|\, \omega_1) = q, \quad \sigma_{\{2,3\}}((O,O) \,|\, \omega_1) = 1 - q,$$

is an equilibrium for $q \geq 1 - {}^2\!/_{5p}$. For $q = 0$, this reduces to the condition $p \leq {}^2\!/_5$ for the second of our pure strategy equilibria above. The expected payoff of individual 2 in this equilibrium is $2q+6(1-q) = 6-4q$ (decreasing in q) and the expected payoff of individual 3 is $6q + 4(1 - q) = 4 + 2q$ (increasing in q). As there is no aggregate payoff function and we deal with Pareto relations, agent $\{2, 3\}$ has no preference ordering over these values of q, but this is not the same as indifference. Finally, we note that

$$\sigma_1(O \,|\, \omega_0) = \sigma_1(O \,|\, \omega_1) = 1, \quad \sigma_2(O \,|\, \omega_0) = 1, \quad \sigma_3(O \,|\, \omega_0) = 1,$$
$$\sigma_{\{2,3\}}((F,F) \,|\, \omega_1) = q, \quad \sigma_{\{2,3\}}((O,O) \,|\, \omega_1) = 1 - q,$$

is never an equilibrium for $q < 1$, the reason being that the expected payoff of individual 2 is $4q + 3(1 - q) = 3 + q$ (increasing in q) and the expected payoff of individual 3 is $6q + 4(1 - q) = 4 + 2q$ (increasing in q), so all $q < 1$ are Pareto dominated by $q = 1$, which corresponds to the first of our pure strategy equilibria above.

REFERENCES

Angus, S.D., Newton, J., 2015. Emergence of shared intentionality is coupled to the advance of cumulative culture. PLoS Comput Biol 11, e1004587.

Arrow, K.J., 1994. Methodological individualism and social knowledge. American Economic Review 84, 1–9.

Aumann, R., 1959. Acceptable points in general cooperative n-person games, in: Tucker, A.W., Luce, R.D. (Eds.), Contributions to the Theory of Games IV. Princeton University Press, pp. 287–324.

Bacharach, M., 1999. Interactive team reasoning: a contribution to the theory of co-operation. Research in economics 53, 117–147.

Bacharach, M., 2006. Beyond individual choice: teams and frames in game theory. Princeton University Press.

Bernheim, B.D., Peleg, B., Whinston, M.D., 1987. Coalition-proof nash equilibria i. concepts. Journal of Economic Theory 42, 1–12.

Cournot, A.A., 1838. Recherches sur les principes mathématiques de la théorie des richesses par Augustin Cournot. chez L. Hachette.

Feldman, A.M., 1974. Recontracting stability. Econometrica 42, pp. 35–44.

Friedman, M., 1953. Essays in positive economics. University of Chicago Press.

Gold, N., Sugden, R., 2007. Collective intentions and team agency. The Journal of Philosophy 104, 109–137.

Green, J.R., 1974. The stability of edgeworth's recontracting process. Econometrica 42, pp. 21–34.

Hardin, G., 1968. The tragedy of the commons. Science 162, 1243–1248. http://science.sciencemag.org/content/162/3859/1243.full.pdf.

Harsanyi, J.C., 1967. Games with incomplete information played by "bayesian" players, i-iii. part i. the basic model. Management Science 14, 159–182.

Harsanyi, J.C., 1968a. Games with incomplete information played by bayesian players part ii. bayesian equilibrium points. Management Science 14, 320–334.

Harsanyi, J.C., 1968b. Games with incomplete information played by'bayesian'players, part iii. the basic probability distribution of the game. Management Science 14, 486–502.

Heath, J., 2015. Methodological individualism, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Spring 2015 edition.

Ichiishi, T., Idzik, A., 1996. Bayesian cooperative choice of strategies. International Journal of Game Theory 25, 455–473.

Kajii, A., Morris, S., 1997. The robustness of equilibria to incomplete information. Econometrica 65, 1283–1309.

Keynes, J.N., 1890. The Scope and Method of Political Economy. Number keynes1890 in History of Economic Thought Books, McMaster University Archive for the History of Economic Thought.

Kuhn, S., 2014. Prisoner's dilemma, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Fall 2014 edition.

Lloyd, W.F., 1833. Two lectures on the checks to population. S. Collingwood.
Moreno, D., Wooders, J., 1996. Coalition-proof equilibrium. Games and Economic Behavior 17, 80–112.
Newton, J., 2012a. Coalitional stochastic stability. Games and Economic Behavior 75, 842–54.
Newton, J., 2012b. Recontracting and stochastic stability in cooperative games. Journal of Economic Theory 147, 364–81.
Nozick, R., 1994. The nature of rationality. Princeton University Press.
Roth, A.E., Vande Vate, J.H., 1990. Random paths to stability in two-sided matching. Econometrica 58, 1475–80.
Samuelson, L., 2016. Game theory in economics and beyond. The Journal of Economic Perspectives 30, 107–130.
Schumpeter, J., 1909. On the concept of social value. The Quarterly Journal of Economics 23, 213–232.
Searle, J., 1990. Collective intentions and actions, in: Cohen, P.R., Morgan, J., Pollack, M. (Eds.), Intentions in communication. MIT Press, pp. 401–15.
Sugden, R., 1993. Thinking as a team: Towards an explanation of nonselfish behavior. Social philosophy and policy 10, 69–89.
Sugden, R., 2003. The logic of team reasoning. Philosophical explorations 6, 165–181.
Tomasello, M., 2014. A natural history of human thinking. Harvard University Press.
Tuomela, R., Miller, K., 1988. We-intentions. Philosophical Studies 53, 367–389.
Weber, M., 1922. Wirtschaft und gesellschaft: Grundriss der verstehenden Soziologie.