Backward induction foundations of the Shapley value

Ben McQuillin and Robert Sugden

July 5, 2016

BACKWARD INDUCTION FOUNDATIONS OF THE SHAPLEY VALUE

Ben McQuillin[1]

and Robert Sugden[2]

We present a noncooperative game model of coalitional bargaining, closely based on that of Gul (1989) but solvable by backward induction. In this game, Gul's condition of 'value additivity' does not suffice to ensure the existence of a subgame perfect Nash equilibrium that supports the Shapley value, but a related condition - 'no positive value-externalities' - does. Multiple equilibria can arise only in the event of ties, and with a mild restriction on tie-break rules these equilibria all support the Shapley value.

Keywords: Shapley value, Nash program, noncooperative coalitional bargaining.

1. INTRODUCTION

This paper builds on the important contribution by Gul (1989, 1999) within a literature of the last three decades that provides 'bargaining foundations' to solution concepts in cooperative game theory. This literature, in the spirit of the 'Nash program', treats the noncooperative and cooperative approaches as mutually illuminative. Gul sets up an intuitively plausible noncooperative bargaining process in which coalitions form through successive pairwise amalgamations (presented as buyouts), and shows that this process is in certain senses supportive of the well known value of Shapley (1953). The process is infinite: coalitional bargaining is allowed to continue for as long as the grand coalition has not formed, and takes place alongside the underlying cooperative event. These aspects of the set-up have certain

[1] b.mcquillin@uea.ac.uk, School of Economics and Centre for Behavioural and Experimental Social Science, University of East Anglia, Norwich NR4 7TJ, UK.

[2] r.sugden@uea.ac.uk, School of Economics and Centre for Behavioural and Experimental Social Science, University of East Anglia, Norwich NR4 7TJ, UK.

1

2

advantages in terms of mathematical tractability, but also certain disadvantages: in particular, where there are subgame perfect Nash equilibria that support the Shapley value there may also be others that do not. In this paper we modify Gul's bargaining process so that it is finite and so that it precedes the underlying cooperative event. The noncooperative game is then solvable by backward induction and we can therefore draw sharper conclusions about the relationship between its subgame perfect Nash equilibria and the Shapley value of the underlying cooperative game.

In Gul's bargaining game, 'value-additivity' ensures that some subgame perfect Nash equilibrium (SPNE) implements the Shapley value, but not that every SPNE does so. In our modified bargaining game it requires a stronger condition - 'no positive value-externalities' - to ensure the existence of a SPNE that implements the Shapley value, but the same condition then ensures that *every* SPNE implements the value.

2. PRELIMINARIES

We consider any *transferable utility game in characteristic function form* (N, v), with N denoting a set of n players, $v \colon 2^N \to \mathbb{R}$ having properties (i) $v(\varnothing) = 0$, and (ii) for any partition π of N, $\sum_{I \in \pi} v(I) \leq v(N)$.

The *Shapley value* assigns to (N, v) an n-tuple, ϕ, of players *expectations* in (N, v), defined by (for any $i \in N$):

$$\phi_i \equiv \sum_{I \subseteq N} \frac{(|I| - 1)!(n - |I|)!}{n!} \left(v(I) - v\left(I \setminus \{i\} \right) \right).$$

Gul (1989) defines a related solution concept: the *generalized Shapley value*. Let Π denote the set of partitions of N, and M the set of *embedded coalitions*, $M \equiv \{(I, \pi) : \pi \in \Pi, I \in \pi\}$. The generalized Shapley value assigns to (N, v) a mapping, $\Phi : M \to \mathbb{R}$, defined by (for any $(I, \pi) \in M$):

$$\Phi(I, \pi) \equiv \sum_{T \subseteq \pi} \frac{(|T| - 1)!(|\pi| - |T|)!}{|\pi|!} \left(v\left(\bigcup_{J \in T} J \right) - v\left(\bigcup_{J \in (T \setminus \{I\})} J \right) \right).$$

For present purposes, it suffices just to view the generalized Shapley value as a formal object; but the intuition is that it assigns expectations to *prior coalitions*, and it does so by treating the prior coalitions as players in a subgame, and by applying the Shapley value to that subgame. Clearly (for any $i \in N$), $\Phi(\{i\}, \{\{j\} : j \in N)\}) = \phi_i$.

We shall define two noncooperative games, based on our underlying co-operative game (N, v), each also with player sets N. The *infinite time bargaining game* is the one set up by Gul, and the *deadline bargaining game* is our modification of Gul's game. In both these games, coalitional bargaining takes place over a sequence of periods. At the start of each period, play-ers are arranged as a partition, or *coalition structure,* each element being a *coalition.* At the start of the first period, the coalition structure is the set of singletons $\{\{i\} : i \in N\}$. At any stage, one player in each coalition is its *representative.* A player is *active* if and only if she represents some coalition. In each period (unless the grand coalition has already formed) there is some opportunity for two coalitions to merge by mutual agreement. Mergers take the form of *buy-outs*: the representative of one of the coalitions makes a payment to the representative of the other, and becomes the sole represen-tative of the merged coalition. Payoffs to embedded coalitions, as specified by v, accrue to the representatives of the corresponding coalitions in the bargaining game. The full definition of each game requires an additional parameter. In the infinite time game, a parameter $\delta \in (0, 1)$ sets a common discount factor; in the deadline game a natural number parameter θ sets the length of the game.

In the infinite time game (with discount factor δ) the sequence of time periods is infinite: $(1, 2, \ldots)$. Players' outcome utilities are given by in-period *payments*, adjusted by δ, so for any player a payment amount of 1 in time period t is worth δ^t units of utility. In each time period, if there are two or more active players the following *in-period bargaining process* occurs ((a)-

(d)). (a) From among the active players, nature selects a pair, with the probability of being selected being equal across active players. (b) Nature then selects one member of this pair to be the *proposer*, with either member being chosen with equal probability and the other member becoming the *responder*. (c) Then the proposer selects an *offer*: any real number. (d) Then the responder decides whether to accept or reject the offer. If she accepts then she receives a payment equal to the offer and ceases to be active. In this case the proposer, correspondingly, makes a payment equal to the offer and becomes the sole representative both of the players she already represented and of the players previously represented by the responder. At the end of each time period, each player that remains active receives a payment amount given by $(1 - \delta)v(A)$ where $A \subseteq N$ is the set of players that she now represents.[1] Then the next time period begins.

In the deadline game (of length θ) the sequence of time periods is finite: $(1, 2, ..., \theta)$. Players' outcome utilities are given directly by total payments, so for any player a payment amount of 1 in any time period is worth one unit of utility. In each period, if there are two or more active players the same in-period bargaining process ((a)-(d) as described above) occurs. In contrast to the infinite time game, there are no additional payments to active players at the end of each time period, except period θ. At the end of period θ, each player that remains active receives a payment given by $v(A)$ where $A \subseteq N$ is the set of players she represents.

Gul's results, and ours, relate to SPNEs in the bargaining games set out above. To simplify analysis, we assume that ties (situations in which the maximum offer a proposer is willing to make exactly equals the minimum offer a responder is willing to accept) within a SPNE lead to an accepted

---

[1] The utility outcome for a player that represents A throughout the bargaining game is therefore $\sum_{t=0}^{\infty} \delta^t (1 - \delta)v(A) = v(A)$.

offer with some fixed exogenous probability, or *tie-break rule*.[2]

Gul uses two further equilibrium conditions: 'stationarity', and that 'every possible meeting ends in agreement'. 'Stationarity' entails that players' strategies specify (for the same coalition structure) the same behavior in every time period. 'Every possible meeting ends in agreement' entails that players' strategies specify, in every situation, offers by proposers that are deemed acceptable by the corresponding responders, so the grand coalition will form directly, in $n-1$ time periods.

3. ANALYSIS OF THE INFINITE TIME BARGAINING GAME

Gul's (1989, 1999) analysis is of the infinite time bargaining game, and his results are obtained in a limit as δ tends to 1.

Gul's first result states that, in any stationary SPNE such that every possible meeting ends in agreement, the expected utility to any player i at the start of the infinite time bargaining game is arbitrarily close (if δ is sufficiently close to 1) to ϕ_i. His second result concerns conditions on v for the existence of such an equilibrium. A necessary condition is *value-additivity* of v, v being value-additive if and only if:

$$\forall \pi \in \Pi, \forall I, J \in \pi, \Phi(I \cup J, (\pi \setminus \{I, J\}) \cup \{I \cup J\}) \geq \Phi(I, \pi) + \Phi(J, \pi).$$

A sufficient condition combines value-additivity with *strict superadditivity*, v being strictly superadditive if and only if for all non-empty and disjoint subsets I and J of N, $v(I \cup J) > v(I) + v(J)$.

---

[2]It is a straightforward matter to also admit to our analysis more elaborate tie-break rules in which the probability of agreement in a tie depends on the time period, the coalition structure, the players represented by the proposer and the players represented by the responder. However, because of the backward induction analysis we use, we cannot admit tie-break rules that depend otherwise on the history of the game or on the specific identities of coalitions' representatives. Gul (1989, 1999) incorporates a similar limitation into his full 'stationarity' condition.

6

So (in the limit, as δ tends to 1) the infinite time game supports the Shapley value, but only provided that v is both superadditive and value-additive, and then only in one SPNE among possible others (which are non-stationary, or in which possible meetings do not end in agreement).[3]

4. ANALYSIS OF THE DEADLINE BARGAINING GAME

Our analysis is of the deadline bargaining game, and our results are obtained in a limit as θ tends to ∞.

The deadline bargaining game preserves Gul's plausible, intuitive procedure of pairwise agreement but, for a given tie-break rule, it has a unique SPNE which can be found by backward induction. Our main result employs a new condition on v, stronger than value-additivity: *no positive value-externalities*, defined as follows. There are no positive value externalities in v if and only if:

$$\forall \pi \in \Pi, \forall I, J, K \in \pi, \Phi(K, \pi) - \Phi(K, (\pi \setminus \{I, J\}) \cup \{I \cup J\}) \geq 0.$$

THEOREM 1 *If there are no positive value-externalities in v then, for any $\varepsilon > 0$, there is an integer t such that, for any $\theta \geqslant t$, in every SPNE of the deadline bargaining game (of length θ), the expected utility of every player i lies within ε of ϕ_i.*

(The proof of Theorem 1 is in the Appendix.)

So, if there are no positive value-externalities in v then (in the limit as θ tends to ∞) the deadline bargaining game supports the Shapley value in *every* SPNE. Relative to the infinite time game, the deadline bargaining

---

[3]Gul (1999) shows that if v is convex then, among all stationary strategy profiles, those in which every possible meeting ends in agreement will have the distinction of being those which are 'efficient' in the sense of maximising total expected utility across players. But Hart and Levy (1999) demonstrate by example that efficiency does not entail immediate agreement more generally.

game supports the Shapley value less equivocally, but on a smaller class of cooperative games. The condition of no positive value-externalities states that within the generalized Shapley value *every* externality associated with any bilateral amalgamation is non-positive, whereas Gul's value-additivity condition can be read as saying that (since the generalized Shapley value summed across any partition is, by construction, equal to the payoff of the grand coalition) the *sum* of value-externalities associated with any bilateral amalgamation is non-positive.[4]

Having observed that a stronger condition than value-additivity ensures, in the relevant limit, that every SPNE in the deadline game supports the Shapley value, we should note the following.

REMARK 1   Value-additivity, even combined with strict superadditivity, is *not* sufficient (unless there are fewer than 5 players) to ensure that there exists *at least one* SPNE of the deadline game that supports the Shapley value.

(The proof of Remark 1 is in the Appendix.)

5. ADDITIONAL OBSERVATIONS

The proof of Theorem 1 leads also to two further results. First, recall that in the infinite time game the condition of value-additivity (plus strict superadditivity) ensures the existence of a stationary SPNE that supports the Shapley value, but fails to ensure the non-existence of a stationary SPNE that does not. We find that the condition of no positive value-externalities

---

[4]Gul's value-additivity condition can be re-written, using the definition of Φ:

$$\forall \pi \in \Pi(N), \forall I, J \in \pi, \sum_{K \in \pi \setminus \{I,J\}} \Phi(K, \pi) - \Phi(K, (\pi \setminus \{I,J\}) \cup \{I \cup J\}) \geq 0.$$

ensures this absence.

REMARK 2   If v has no positive value-externalities then, in the relevant limit, the infinite time game supports the Shapley value in *every* stationary SPNE.

Second, we have so far presented an analysis based on an underlying cooperative game in characteristic function form. However, the structure of our proofs is such that we could as easily have presented the same analysis based on an underlying game (N, w) in *partition function form*, $w : M \to \mathbb{R}$. (In a deadline bargaining game based on (N, w), at the end of period θ, the game would end with each player that remains active receiving a payment given by $w(A, \pi)$ where $A \subseteq N$ is the set of players she represents and π is the prevailing coalition structure.) For a partition function $w : M \to \mathbb{R}$, we define its *characteristic function reduction* $\mathcal{T}(w) \colon 2^N \to \mathbb{R}$, using $\forall I \subseteq N, \mathcal{T}(w)(I) \equiv w(I, \{I, N \setminus I\})$. The *extended Shapley value*, characterized in McQuillin (2009), assigns to (N, w) an n-tuple of expectations which is just the Shapley value of $(N, \mathcal{T}(w))$. And the following result holds.

REMARK 3   If there are no positive value-externalities in the characteristic function reduction of w then (in the relevant limits) the *extended* Shapley value is supported both by the deadline game (in every SPNE) and by the infinite time game (in every stationary SPNE).

(The proofs of Remarks 2 and 3 are in the Appendix.)

6. DISCUSSION

Our paper is a contribution to a well-established project of game theory. The aim of this project is to build stylized models of real-world bargaining procedures and to investigate the conditions under which solution concepts from cooperative game theory are induced by fully rational bargaining. Our

model is based on that of Gul (1989), and retains its key feature, which is that cooperative structures arise through sequences of bilateral agreements. In contrast, in alternative implementations of the Shapley value such as proposed by Hart and Mas-Colell (1996) and Pérez-Castrillo and Wettstein (2001) the coalition structure changes only once, and there is never more than one non-singleton coalition. Which of these modelling strategies is more useful depends on the bargaining procedure that is to be represented. However, many institutional frameworks for real-world bargaining, most notably the merger and takeover process by which firms amalgamate[5], involve sequences of coalition formation similar to those of our model. It is therefore striking to find that the Shapley value - a concept that is often interpreted as based on normative axioms - can emerge as the outcome in a noncooperative model of bargaining that is applicable to such institutions.[6] Whether the Shapley value does so emerge depends on properties of the underlying cooperative game.

Gul (1989, p. 90) suggests that value-additivity is "key in determining whether a given characteristic function game constitutes a suitable framework for the application of the Shapley value". Our results in this paper suggest rather that a closely related, but stronger condition is key in this sense. Value-additivity (combined with strict superadditivity) ensures that Gul's infinite time bargaining game has a stationary SPNE that supports the Shapley value but there may be other stationary SPNEs that do not. In our deadline bargaining game, value-additivity and strict superadditivity do not guarantee the existence of a SPNE that supports the Shapley value. However, the condition of no positive value-externalities ensures that every stationary SPNE in Gul's game and every SPNE in our game supports the

---

[5]This point is developed in Macho-Stadler et al (2006).

[6]The Shapley value emerges as the *expectatation* within the bargaining game, which is consistent with the way that Shapley (1953) conceived the value.

Shapley value.

APPENDIX: PROOFS

Our proofs are constructed using the set Γ of *games in partition function form on the player set N*, generically $w\colon M \to \mathbb{R}$ such that, for any $\pi \in \Pi$, $w(\varnothing, \pi) = 0$ and $\sum_{I \in \pi} w(I, \pi) \leq w(N, \{N\})$. Notice that in any SPNE of the deadline bargaining game, with any given number of time periods remaining, for each embedded coalition (I, π), there is a well-defined *expectation*: i.e. the expected value of future in-period payments to the player who represents I. The function that assigns these expectations to embedded coalitions is an element of Γ. Also, Φ is an element of Γ.

In the deadline game, each SPNE corresponds to a tie-break rule: the exogenously given probability with which, when the maximum offer a proposer is willing to make exactly equals the minimum offer a responder is willing to accept, this offer is accepted. We denote this probability $\rho \in [0, 1]$. For a given SPNE, the relationship between expectations in adjacent time periods can be described using formal constructions that we shall term *backward induction sequences*. The items within these sequences are elements of Γ.

In this appendix we shall first define a backward induction sequence and then set down a lemma in respect of this construction. Theorem 1 and also Remarks 2 and 3 follow almost immediately from the lemma. Remark 1 is proved by an example.

Given $w \in \Gamma$, we defined a characteristic function $\mathcal{T}(w)\colon 2^N \to \mathbb{R}$ using $\forall I \subseteq N, \mathcal{T}(w)(I) \equiv w(I, \{I, N \setminus I\})$. From the characteristic function $v\colon 2^N \to \mathbb{R}$, we define $\mathcal{S}(v) \in \Gamma$ using: $\forall (I, \pi) \in M, \mathcal{S}(v)(I, \pi) \equiv v(I)$.

We shall from this point use $\Phi(v)$, where we have previously used Φ, to denote the generalized Shapley value of (N, v). This more general notation allows us to reference the generalized Shapley value of *other* transferable

utility characteristic function games (on N), for example the generalized Shapley value of $(N, \acute{v})$ becomes $\Phi(\acute{v})$.

A.1. *Backward induction sequences*

We need to set up some additional notation: terms that describe the internal and external effects, within some $w \in \Gamma$, of coalitions amalgamating. Given $\pi \in \Pi$, we shall use $\overset{IJ}{\pi}$ as a shorthand for $(\pi \setminus \{I, J\}) \cup \{I \cup J\}$. For any $(A, \pi) \in M$, for any pair of coalitions $\{I, J\} \subseteq \pi$, and for any $\rho \in [0, 1]$, we define two formal objects, $F_w^{\{I,J\}}(A, \pi)$ and $G_{w,\rho}^{\{I,J\}}(A, \pi)$. The first term can be read as 'the effect on A if I and J amalgamate'; the second term can be read as 'the effect on A if I and J amalgamate if the interior effects are positive, do not do so if the interior effects are negative, and do so with probability ρ if the interior effects are null'. In either case the supposition is that A may or may not be a member of $\{I, J\}$, and that amalgamating coalitions share the surplus equally. So:

$$F_w^{\{I,J\}}(A, \pi) \equiv \begin{cases} \frac{w\left(I \cup J, \overset{IJ}{\pi}\right) - w(I, \pi) - w(J, \pi)}{2} & \text{where } A \in \{I, J\} \\ w\left(A, \overset{IJ}{\pi}\right) - w(A, \pi) & \text{where } A \notin \{I, J\} \end{cases} .$$

$$G_{w,\rho}^{\{I,J\}}(A, \pi) \equiv \begin{cases} F_w^{\{I,J\}}(A, \pi) & \text{where } w\left(I \cup J, \overset{IJ}{\pi}\right) > w(I, \pi) + w(J, \pi) \\ \rho F_w^{\{I,J\}}(A, \pi) & \text{where } w\left(I \cup J, \overset{IJ}{\pi}\right) = w(I, \pi) + w(J, \pi) \\ 0 & \text{where } w\left(I \cup J, \overset{IJ}{\pi}\right) < w(I, \pi) + w(J, \pi) \end{cases} .$$

We now consider the deadline bargaining game (of length θ) defined on v. Suppose, in the SPNE associated with the tie-break rule ρ, with $t < \theta$ periods remaining in this game, expected future payments to the remaining active players are described by $w_t \in \Gamma$. That is, the player that represents coalition I within the coalition structure π expects a sum of payments over periods $((\theta - t), (\theta - t + 1), ..., \theta)$ equal to $w_t(I, \pi)$. By backward induction, in the same equilibrium of the bargaining game, with $t + 1$ periods remaining,

expected future payments to the remaining active players are described by $w_{t+1} \in \Gamma$, satisfying:

$$\forall (A, \pi) \in M, w_{t+1}(A, \pi) = w_t(A, \pi) + \frac{2 \sum_{\{I,J\} \subseteq \pi} G_{w_t,\rho}^{\{I,J\}}(A, \pi)}{|\pi| \, (|\pi| - 1)}. \quad (1)$$

We define a *backward induction sequence* to be a sequence $\{w_t\}_{t=0}^{\infty}$ in Γ with the property that there exists $\rho \in [0, 1]$ such that, for any non-negative integer t, equation (1) is satisfied.

A.2. *Lemma 1*

We can now state our Lemma.

LEMMA 1 *For any $w \in \Gamma$, if $\mathcal{T}(w)$ has no positive value-externalities, then for any $\varepsilon > 0$, there exists $t' \in \mathbb{N}$ such that, for any $t \geqslant t'$, for any $(I, \pi) \in M$, in any backward induction sequence $\{w_t\}_{t=0}^{\infty}$ with $w_0 = w$, $|w_t(I, \pi) - \Phi(\mathcal{T}(w))(I, \pi)| < \varepsilon$.*

A.3. *Proof of Lemma 1*

For any $w \in \Gamma$ we define $\tilde{w} \equiv w - \Phi(\mathcal{T}(w))$. $\Phi(\mathcal{T}(w))$ has, by construction, the following 'efficiency' property:

$$\forall \pi \in \Pi, \sum_{I \in \pi} \Phi(\mathcal{T}(w))(I, \pi) = w(N, \{N\}). \quad (2)$$

And it also has, using results in McQuillin (2009) (Theorems 3 and 2 respectively), the following two properties:

$$\forall (A, \pi) \in M, \sum_{\{I,J\} \subseteq \pi} F_{\Phi(\mathcal{T}(w))}^{\{I,J\}}(A, \pi) = 0. \quad (3)$$

$$\Phi(\mathcal{T}(\Phi(\mathcal{T}(w)))) = \Phi(\mathcal{T}(w)). \quad (4)$$

Any backward induction sequence $\{w_t\}_{t=0}^{\infty}$ has the following properties:

$$\forall t \in \mathbb{Z}_{\geqslant 0}, \forall \pi \in \Pi, \sum_{I \in \pi} \widetilde{w_t}(I, \pi) \leqslant 0. \tag{5}$$

$$\forall t \in \mathbb{Z}_{\geqslant 1}, \forall (A, \pi) \in M, |\pi| \leqslant 2 \rightarrow w_t(A, \pi) - \Phi(\mathcal{T}(w_0)) = 0. \tag{6}$$

And (using (4) and (6)):

$$\forall t \in \mathbb{Z}_{\geqslant 0}, \Phi(\mathcal{T}(w_t)) = \Phi(\mathcal{T}(w_0))$$

$$\therefore \quad \widetilde{w_t} = w_t - \Phi(\mathcal{T}(w_0)). \tag{7}$$

We now consider some partition function game $w \colon M \to \mathbb{R}$ such that $\mathcal{T}(w)$ has no positive value-externalities, and we consider the set of all backward induction sequences $\{w_t\}_{t=0}^{\infty}$ with $w_0 = w$. We write this set as $\{\{w_{\rho,t}\}_{t=0}^{\infty} : \rho \in [0, 1], w_{\rho,0} = w\}$, with:

$$\forall t \in \mathbb{Z}_{\geqslant 0}, \forall (A, \pi) \in M, \forall \rho \in [0, 1],$$

$$w_{\rho, t+1}(A, \pi) = w_{\rho, t}(A, \pi) + \frac{2 \sum_{\{I,J\} \subseteq \pi} G_{w_t, \rho}^{\{I,J\}}(A, \pi)}{|\pi| (|\pi| - 1)}$$

$$\therefore \quad \widetilde{w_{\rho, t+1}}(A, \pi) = \widetilde{w_{\rho, t}}(A, \pi) + \frac{2 \sum_{\{I,J\} \subseteq \pi} G_{w_t, \rho}^{\{I,J\}}(A, \pi)}{|\pi| (|\pi| - 1)}. \tag{8}$$

Using (2) and (3):

$$\forall (A, \pi) \in M, \forall \rho \in [0, 1], \sum_{\{I,J\} \subseteq \pi} G_{\Phi(\mathcal{T}(w)), \rho}^{\{I,J\}}(A, \pi) = 0. \tag{9}$$

We proceed by induction. (We simplify the induction hypothesis by disregarding epsilons.) Suppose, for some positive integers $k \geqslant 2$ and t':

$$\forall t \in \mathbb{Z}_{\geqslant t'}, \forall \rho \in [0, 1], \forall (A, \pi) \in M, |\pi| \leqslant k \rightarrow \widetilde{w_{\rho, t}}(A, \pi) = 0. \tag{10}$$

((6) and (7) establish this hypothesis for $k = 2$.) Our aim is to show that then:

$$\forall \varepsilon > 0, \exists t'' \in \mathbb{Z}_{\geqslant t'}, \forall \pi \in \Pi, \forall t \in \mathbb{Z}_{\geqslant t''}, \forall \rho \in [0, 1], \forall (A, \pi) \in M,$$

$$|\pi| \leqslant k + 1 \rightarrow |\widetilde{w_{\rho, t}}(A, \pi)| < \varepsilon. \tag{11}$$

Consider any $\pi \in \Pi$, such that $|\pi| = k+1$, and any integer $t > t'$. Then consider any $\rho \in [0,1]$ and any $A \in \pi$ such that $\widetilde{w_{\rho,t}}(A,\pi) \leqslant 0$. Using the induction hypothesis (10):

$$\forall I \ \in \ \pi \setminus \{A\}, G^{\{I,A\}}_{w_{\rho,t}}(A,\pi) \geqslant G^{\{I,A\}}_{\Phi(\mathcal{T}(w))}(A,\pi) - \frac{(\widetilde{w_{\rho,t}}(A,\pi) + \widetilde{w_{\rho,t}}(I,\pi))}{2}$$

$$\therefore \sum_{I \in \pi \setminus \{A\}} G^{\{I,A\}}_{w_{\rho,t}}(A,\pi) \geqslant \left(\sum_{I \in \pi \setminus \{A\}} G^{\{I,A\}}_{\Phi(\mathcal{T}(w))}(A,\pi) \right) - \left(\sum_{I \in \pi \setminus \{A\}} \frac{\widetilde{w_{\rho,t}}(I,\pi)}{2} \right)$$
$$- \frac{(|\pi| - 1)\widetilde{w_{\rho,t}}(A,\pi)}{2}$$

$$\therefore \sum_{I \in \pi \setminus \{A\}} G^{\{I,A\}}_{w_{\rho,t}}(A,\pi) \geqslant \left(\sum_{I \in \pi \setminus \{A\}} G^{\{I,A\}}_{\Phi(\mathcal{T}(w))}(A,\pi) \right) - \left(\sum_{I \in \pi} \frac{\widetilde{w_{\rho,t}}(I,\pi)}{2} \right)$$
$$- \frac{(|\pi| - 2)\widetilde{w_{\rho,t}}(A,\pi)}{2}. \tag{12}$$

Combining (5) and (12) gives:

$$\sum_{I \in \pi \setminus \{A\}} G^{\{I,A\}}_{w_{\rho,t}}(A,\pi) \geqslant \left(\sum_{I \in \pi \setminus \{A\}} G^{\{I,A\}}_{\Phi(\mathcal{T}(w))}(A,\pi) \right) - \frac{(|\pi| - 2)\widetilde{w_{\rho,t}}(A,\pi)}{2}. \tag{13}$$

Note also (again using the induction hypothesis, and as there are no positive value-externalities in $\mathcal{T}(w)$):

$$\forall \{I,J\} \in \pi \setminus \{A\}, G^{\{I,J\}}_{w_{\rho,t}}(A,\pi) \geqslant G^{\{I,J\}}_{\Phi(\mathcal{T}(w))}(A,\pi). \tag{14}$$

Combining (13) and (14) gives:

$$\sum_{\{I,J\} \subseteq \pi} G^{\{I,J\}}_{w_{\rho,t}}(A,\pi) \geqslant \left(\sum_{\{I,J\} \subseteq \pi} G^{\{I,J\}}_{\Phi(\mathcal{T}(w))}(A,\pi) \right) - \frac{(|\pi| - 2)\widetilde{w_{\rho,t}}(A,\pi)}{2}. \tag{15}$$

Combining (9) and (15) gives:

$$\sum_{\{I,J\} \subseteq \pi} G^{\{I,J\}}_{w_{\rho,t}}(A,\pi) \geqslant \frac{-(|\pi| - 2)\widetilde{w_{\rho,t}}(A,\pi)}{2}. \tag{16}$$

Combining (8) and (16) gives:
$$\widetilde{w_{\rho,t+1}}(A,\pi) \geqslant \frac{(-(|\pi| - 2)\widetilde{w_{\rho,t}}(A,\pi))}{|\pi|\,(|\pi| - 1)} + \widetilde{w_{\rho,t}}(A,\pi)$$
$$\therefore \widetilde{w_{\rho,t+1}}(A,\pi) \geqslant \left(1 - \frac{(|\pi| - 2)}{|\pi|\,(|\pi| - 1)} \right) \widetilde{w_{\rho,t}}(A,\pi). \tag{17}$$

Now consider any $B \in \pi$ such that $\widetilde{w_{\rho,t}}(B, \pi) > 0$. Note (recollecting the steps that led to (13)):

$$\sum_{I \in \pi \setminus \{B\}} G_{w_{\rho,t}}^{\{I,B\}}(A, \pi) > \left(\sum_{I \in \pi \setminus \{B\}} G_{\Phi(\mathcal{T}(w))}^{\{I,B\}}(B, \pi) \right) - \frac{(|\pi| - 2)\widetilde{w_{\rho,t}}(B, \pi)}{2}. \quad (18)$$

Note also:

$$\forall \{I, J\} \in \pi \setminus \{B\}, G_{w_{\rho,t}}^{\{I,J\}}(B, \pi) \geqslant G_{\Phi(\mathcal{T}(w))}^{\{I,J\}}(B, \pi) - \widetilde{w_{\rho,t}}(B, \pi). \quad (19)$$

Combining (9), (18) and (19) gives:

$$\sum_{\{I,J\} \subseteq \pi} G_{w_{\rho,t}}^{\{I,J\}}(A, \pi) > -\frac{|\pi|\,(|\pi| - 2)\widetilde{w_{\rho,t}}(B, \pi)}{2}. \quad (20)$$

Combining (8) and (20), and collecting terms gives:
$$\begin{aligned} \widetilde{w_{\rho,t+1}}(B, \pi) &> \frac{\widetilde{w_{\rho,t}}(B, \pi)}{(|\pi| - 1)} \\ \therefore \widetilde{w_{\rho,t+1}}(B, \pi) &> 0. \end{aligned} \quad (21)$$

It is clear from (5), (17) and (21) that:

$$0 \geqslant \min_{\rho \in [0,1], A \in \pi} \widetilde{w_{\rho,t+1}}(A, \pi) \geqslant \left(1 - \frac{(|\pi| - 2)}{|\pi|\,(|\pi| - 1)} \right) \min_{\rho \in [0,1], A \in \pi} \widetilde{w_{\rho,t}}(A, \pi). \quad (22)$$

And it is clear from (5) and (22) that $\left\{ \max_{\rho \in [0,1], A \in \pi} |\widetilde{w_{\rho,t}}(A, \pi)| \right\}_{t=0}^{\infty}$ converges to zero. This gives us our proof of (11) and therefore of Lemma 1.

A.4. *Proof of Theorem 1*

In any SPNE of the deadline bargaining game (of length θ), the expected utility to player i is given by the term $w_{\theta}(\{i\}, \{\{j\} : j \in N)\})$ in a backward induction sequence $\{w_t\}_{t=0}^{\infty}$ with $w_0 = \mathcal{S}(v)$. So, by Lemma 1 we directly have Theorem 1.

A.5. *Proof of Remark 1*

Remark 1 is established by the following example. We suppose $N = \{1, 2, 3, 4, 5\}$. (It is easy to show that in games with fewer than four players, value-additivity implies no positive value-externalities. One can also show that in games with four players value-additivity suffices to ensure that the deadline game supports the Shapley value, but in this case the proof is cumbersome and unenlightening.) We then define $v \equiv (v_a + v_b + v_c + v_d) \times 12 + \frac{v_e}{10000}$ where:

$$v_a(A) = \begin{cases} 1 & \text{where } \{1, 2, 4\} \subseteq A \\ 0 & \text{otherwise} \end{cases}$$

$$v_b(A) = \begin{cases} 1 & \text{where } |\{1, 2, 4, 5\} \cap A| \geqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

$$v_c(A) = \begin{cases} 1 & \text{where } \{1, 3, 5\} \subseteq A \\ 0 & \text{otherwise} \end{cases}$$

$$v_d(A) = \begin{cases} 1 & \text{where } |\{1, 3, 4, 5\} \cap A| \geqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

$$v_e(A) = |A| - 1.$$

We show $\Phi(v')$ with $v' \equiv (v_a + v_b + v_c + v_d) \times 12$ in Figure 1. Then $\Phi(v_e)$ is given by $\Phi(v_e)(A, \pi) = \left(|A| - \frac{1}{|\pi|}\right)$, and we have $\Phi(v) = \Phi(v') + \frac{1}{10000}\Phi(v_e)$. It should be noted that v' is value-additive and weakly superadditive, and v_e value-additive and strictly superadditive; so v is value-additive and strictly superadditive. In a backward induction sequence, $\Phi(v)$ is a stationary point: that is, for $w^* \equiv \Phi(v)$, for any $\rho \in [0, 1]$, for any $(A, \pi) \in M$:

$$w^*(A, \pi) = w^*(A, \pi) + \frac{2 \sum_{\{I, J\} \subseteq \pi} G_{w^*, \rho}^{\{I, J\}}(A, \pi)}{|\pi|(|\pi| - 1)}.$$

However, the same stationarity property holds for w^* defined, with $\dot\pi \equiv$

$\{\{j\} : j \in N\}$, as follows:

$$w^*\left(\{1\}, \dot{\pi}\right) \;=\; \Phi(v)\left(\{1\}, \dot{\pi}\right) - \tfrac{44}{99} + \tfrac{22}{660}\tfrac{1}{10000}$$

$$w^*\left(\{2\}, \dot{\pi}\right) \;=\; w^*\left(\{3\}, \dot{\pi}\right) \;=\; \Phi(v)\left(\{2\}, \pi\right) + \tfrac{2}{99} - \tfrac{13}{660}\tfrac{1}{10000}$$

$$w^*\left(\{4\}, \dot{\pi}\right) \;=\; w^*\left(\{5\}, \pi\right) \;=\; \Phi(v)\left(\{4\}, \pi\right) + \tfrac{20}{99} + \tfrac{2}{660}\tfrac{1}{10000}$$

$$w^*\left(A, \pi\right) \;=\; \Phi(v)\left(A, \pi\right) \quad \text{where } \pi \neq \dot{\pi}.$$

[**Figure 1 around here.**]

Both of these stationary points correspond to expectations within stationary SPNEs in the infinite time bargaining game. (The first corresponds to a stationary SPNE in which all meetings end in agreement; the second to a stationary SPNE in which meetings end in agreement *except*, while the singleton structure prevails, meetings between $\{2\}$ and $\{3\}$, between $\{2\}$ and $\{4\}$ or between $\{3\}$ and $\{5\}$.) But it transpires that it is to the second of these stationary points, which is not $\Phi(v)$, that the (unique) backward induction sequence $\{w_t\}_{t=0}^{\infty}$ with $w_0 \equiv \mathcal{S}(v)$ converges. The graphs in Figures 2 and 3 plot the associated sequence $\{\widetilde{w_t}\}_{t=0}^{\infty}$, $\widetilde{w_t} \equiv w_t - \Phi(v)$, for coalitions embedded in the finest coalition structure (Figure 2), and in coalition structures of cardinality four (Figure 3). (Figure 3 shows six of the ten coalition structures of cardinality four, the remainder, entailing coalitions $\{1,3\}$, $\{1,5\}$, $\{3,4\}$ and $\{3,5\}$, follow by symmetries with those shown entailing $\{1,2\}$, $\{1,4\}$, $\{2,5\}$ and $\{2,4\}$ respectively.) Figure 2 illustrates, for example, that the sequence $\{\widetilde{w_t}(\{1\}, \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\})\}_{t=0}^{\infty}$ converges to $\left(-\tfrac{44}{99} + \tfrac{22}{660}\tfrac{1}{10000}\right)$ from above.

[**Figure 2 around here.**]

[**Figure 3 around here.**]

18

A.6. *Proof of Remark 2*

Gul (1989) shows that, in the limit as δ tends to 1, in every stationary SPNE of the infinite time bargaining game defined on v, at the start of any time period, expected future payments to the remaining active players are described by some $w^* \in \Gamma$ that is a stationary point in a backward induction sequence and that fulfils the following:

$$\forall (I, \pi) \in M, |\pi| \leqslant 2 \rightarrow w^*(I, \pi) = \Phi(v)(I, \pi). \tag{23}$$

If w^* fulfils equation (23) then $\mathcal{T}(w^*)$ equals $\mathcal{T}(\Phi(v))$, and $\Phi(\mathcal{T}(w^*))$ equals $\Phi(\mathcal{T}(\Phi(v)))$. McQuillin (2009) shows that $\Phi(\mathcal{T}(\Phi(v)))$ equals $\Phi(v)$, so therefore $\Phi(\mathcal{T}(w^*))$ equals $\Phi(v)$; and if v has no positive value-externalities then $\mathcal{T}(w^*)$ has no positive value-externalities. By Lemma 1, if w^* is a stationary point in a backward induction sequence it must be that w^* equals $\Phi(\mathcal{T}(w^*))$ and therefore w^* equals $\Phi(v)$ and so $\forall i \in N, w^*(\{i\}, \{\{j\} : j \in N)\}) = \phi_i$.

A.7. *Proof of Remark 3*

In any SPNE of the deadline bargaining game (of length θ) defined on (N, w), the expected utility to player i is given by the term $w_\theta(\{i\}, \{\{j\} : j \in N)\})$ in a backward induction sequence $\{w_t\}_{t=0}^{\infty}$ with $w_0 = w$. So, by Lemma 1 we directly have the result that (as θ tends to ∞) the extended Shapley value is supported by the deadline game (in every SPNE). Analogous reasoning to that used in the proof of Remark 2 gives the result that (as δ tends to 1) the extended Shapley value is supported by the infinite time game (in every stationary SPNE).

of the UK (award no. ES/K002201/1).

REFERENCES

GUL, F. (1989) Bargaining foundations of Shapley value, *Econometrica* **57**, 81-95.

——— (1999) Efficiency and immediate agreement: A reply to Hart and Levy, *Econometrica* **67**, 913-917.

HART, S. AND Z. LEVY (1999) Efficiency does not imply immediate agreement, *Econometrica* **67**, 909-912.

HART, S. AND A. MAS-COLELL (1996) Bargaining and value, *Econometrica* **64**, 357-380.

MACHO-STADLER, I., D. PÉREZ-CASTRILLO AND N. PORTEIRO (2006) Sequential formation of coalitions through bilateral agreements in a Cournot setting, *International Journal of Game Theory* **34**, 207-228.

MCQUILLIN, B. (2009) The extended and generalized Shapley value: Simultaneous consideration of coalitional externalities and coalitional structure, *Journal of Economic Theory* **144**, 696-721.

PÉREZ-CASTRILLO, D. AND D. WETTSTEIN (2001) Bidding for the surplus: A noncooperative approach to the Shapley value, *Journal of Economic Theory* **100**, 274–294.

SHAPLEY, L.S. (1953) A value for n-person games, in *Contributions to the Theory of Games II* (H.W. Kuhn and H.W. Tucker eds.), Princeton University Press.

A	π	$\Phi(v')(A,\pi)$
{1}	{{1},{2},{3},{4},{5}}	14
	{{1},{2},{3},{4,5}}	12
	{{1},{2},{3,4},{5}}	13
	{{1},{2},{3,5},{4}}	15
	{{1},{2},{3,4,5}}	12
	{{1},{2,3},{4},{5}}	14
	{{1},{2,3},{4,5}}	12
	{{1},{2,4},{3},{5}}	15
	{{1},{2,4},{3,5}}	16
	{{1},{2,3,4},{5}}	14
	{{1},{2,5},{3},{4}}	13
	{{1},{2,5},{3,4}}	12
	{{1},{2,3,5},{4}}	14
	{{1},{2,4,5},{3}}	12
	{{1},{2,3,4,5}}	12
{2}	{{1},{2},{3},{4},{5}}	7
	{{1},{2},{3},{4,5}}	6
	{{1},{2},{3,4},{5}}	7
	{{1},{2},{3,5},{4}}	7
	{{1},{2},{3,4,5}}	6
	{{1,3},{2},{4},{5}}	7
	{{1,3},{2},{4,5}}	6
	{{1,4},{2},{3},{5}}	8
	{{1,4},{2},{3,5}}	8
	{{1,3,4},{2},{5}}	8
	{{1,5},{2},{3},{4}}	6
	{{1,5},{2},{3,4}}	6
	{{1,3,5},{2},{4}}	6
	{{1,4,5},{2},{3}}	6
	{{1,3,4,5},{2}}	6
{3}	{{1},{2},{3},{4},{5}}	7
	{{1},{2},{3},{4,5}}	6
	{{1},{2,4},{3},{5}}	7
	{{1},{2,5},{3},{4}}	7
	{{1},{2,4,5},{3}}	6
	{{1,2},{3},{4},{5}}	7
	{{1,2},{3},{4,5}}	6
	{{1,4},{2},{3},{5}}	6
	{{1,4},{2,5},{3}}	6
	{{1,2,4},{3},{5}}	6
	{{1,5},{2},{3},{4}}	8
	{{1,5},{2,4},{3}}	8
	{{1,2,5},{3},{4}}	8
	{{1,4,5},{2},{3}}	6
	{{1,2,4,5},{3}}	6
{4}	{{1},{2},{3},{4},{5}}	10
	{{1},{2},{3,5},{4}}	9
	{{1},{2,3},{4},{5}}	10
	{{1},{2,5},{3},{4}}	9
	{{1},{2,3,5},{4}}	8
	{{1,2},{3},{4},{5}}	11

A	π	$\Phi(v')(A,\pi)$
	{{1,2},{3,5},{4}}	10
	{{1,3},{2},{4},{5}}	9
	{{1,3},{2,5},{4}}	8
	{{1,2,3},{4},{5}}	10
	{{1,5},{2},{3},{4}}	8
	{{1,5},{2,3},{4}}	8
	{{1,2,5},{3},{4}}	8
	{{1,3,5},{2},{4}}	6
	{{1,2,3,5},{4}}	6
{5}	{{1},{2},{3},{4},{5}}	10
	{{1},{2},{3,4},{5}}	9
	{{1},{2,3},{4},{5}}	10
	{{1},{2,4},{3},{5}}	9
	{{1},{2,3,4},{5}}	8
	{{1,2},{3},{4},{5}}	9
	{{1,2},{3,4},{5}}	8
	{{1,3},{2},{4},{5}}	11
	{{1,3},{2,4},{5}}	10
	{{1,2,3},{4},{5}}	10
	{{1,4},{2},{3},{5}}	8
	{{1,4},{2,3},{5}}	8
	{{1,2,4},{3},{5}}	6
	{{1,3,4},{2},{5}}	8
	{{1,2,3,4},{5}}	6
{1,2}	{{1,2},{3},{4},{5}}	21
	{{1,2},{3},{4,5}}	18
	{{1,2},{3,4},{5}}	20
	{{1,2},{3,5},{4}}	22
	{{1,2},{3,4,5}}	18
{1,3}	{{1,3},{2},{4},{5}}	21
	{{1,3},{2},{4,5}}	18
	{{1,3},{2,4},{5}}	22
	{{1,3},{2,5},{4}}	20
	{{1,3},{2,4,5}}	18
{1,4}	{{1,4},{2},{3},{5}}	26
	{{1,4},{2},{3,5}}	26
	{{1,4},{2,3},{5}}	26
	{{1,4},{2,5},{3}}	24
	{{1,4},{2,3,5}}	24
{1,5}	{{1,5},{2},{3},{4}}	26
	{{1,5},{2},{3,4}}	24
	{{1,5},{2,3},{4}}	26
	{{1,5},{2,4},{3}}	26
	{{1,5},{2,3,4}}	24
{2,3}	{{1},{2,3},{4},{5}}	14
	{{1},{2,3},{4,5}}	12
	{{1,4},{2,3},{5}}	14
	{{1,5},{2,3},{4}}	14
	{{1,4,5},{2,3}}	12
{2,4}	{{1},{2,4},{3},{5}}	17
	{{1},{2,4},{3,5}}	16

A	π	$\Phi(v')(A,\pi)$
	{{1,3},{2,4},{5}}	16
	{{1,5},{2,4},{3}}	14
	{{1,3,5},{2,4}}	12
{2,5}	{{1},{2,5},{3},{4}}	19
	{{1},{2,5},{3,4}}	18
	{{1,3},{2,5},{4}}	20
	{{1,4},{2,5},{3}}	18
	{{1,3,4},{2,5}}	18
{3,4}	{{1},{2},{3,4},{5}}	19
	{{1},{2,5},{3,4}}	18
	{{1,2},{3,4},{5}}	20
	{{1,5},{2},{3,4}}	18
	{{1,2,5},{3,4}}	18
{3,5}	{{1},{2},{3,5},{4}}	17
	{{1},{2,4},{3,5}}	16
	{{1,2},{3,5},{4}}	16
	{{1,4},{2},{3,5}}	14
	{{1,2,4},{3,5}}	12
{4,5}	{{1},{2},{3},{4,5}}	24
	{{1},{2,3},{4,5}}	24
	{{1,2},{3},{4,5}}	24
	{{1,3},{2},{4,5}}	24
	{{1,2,3},{4,5}}	24
{1,2,3}	{{1,2,3},{4},{5}}	28
	{{1,2,3},{4,5}}	24
{1,2,4}	{{1,2,4},{3},{5}}	36
	{{1,2,4},{3,5}}	36
{1,2,5}	{{1,2,5},{3},{4}}	32
	{{1,2,5},{3,4}}	30
{1,3,4}	{{1,3,4},{2},{5}}	32
	{{1,3,4},{2,5}}	30
{1,3,5}	{{1,3,5},{2},{4}}	36
	{{1,3,5},{2,4}}	36
{1,4,5}	{{1,4,5},{2},{3}}	36
	{{1,4,5},{2,3}}	36
{2,3,4}	{{1},{2,3,4},{5}}	26
	{{1,5},{2,3,4}}	24
{2,3,5}	{{1},{2,3,5},{4}}	26
	{{1,4},{2,3,5}}	24
{2,4,5}	{{1},{2,4,5},{3}}	30
	{{1,3},{2,4,5}}	30
{3,4,5}	{{1},{2},{3,4,5}}	30
	{{1,2},{3,4,5}}	30
{1,2,3,4}	{{1,2,3,4},{5}}	42
{1,2,3,5}	{{1,2,3,5},{4}}	42
{1,2,4,5}	{{1,2,4,5},{3}}	42
{1,3,4,5}	{{1,3,4,5},{2}}	42
{2,3,4,5}	{{1},{2,3,4,5}}	36
{1,2,3,4,5}	{{1,2,3,4,5}}	48

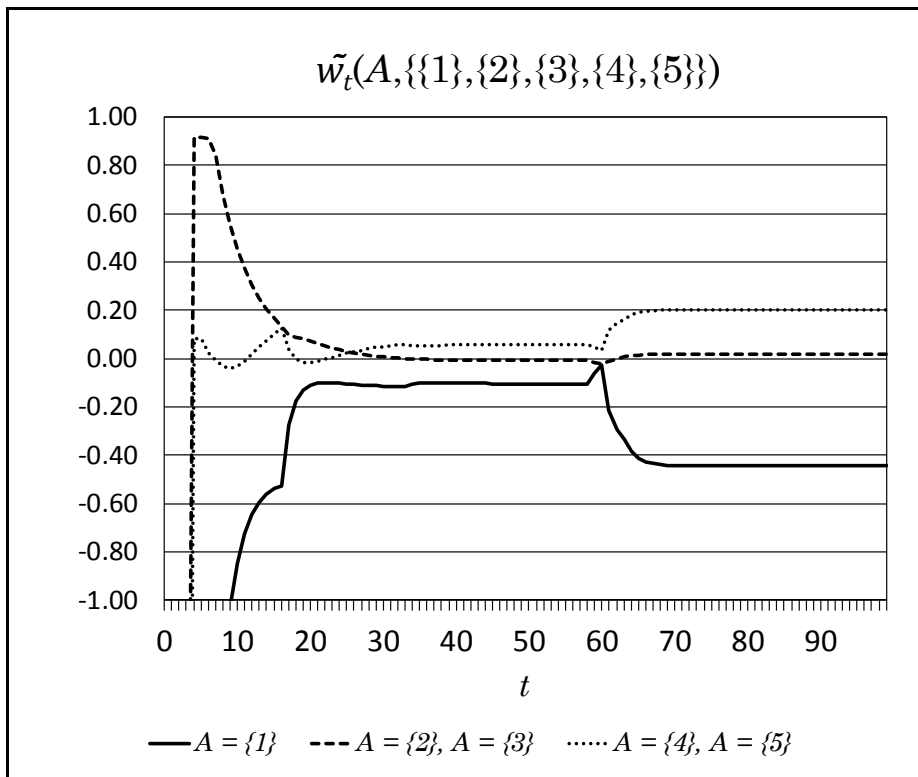Figure 1: $\Phi(v')$ with $v' \equiv (v_a + v_b + v_c + v_d) \times 12$.

Figure 2: $\widetilde{w_t}(A, \pi) \equiv w_t(A, \pi) - \Phi(v)(A, \pi)$
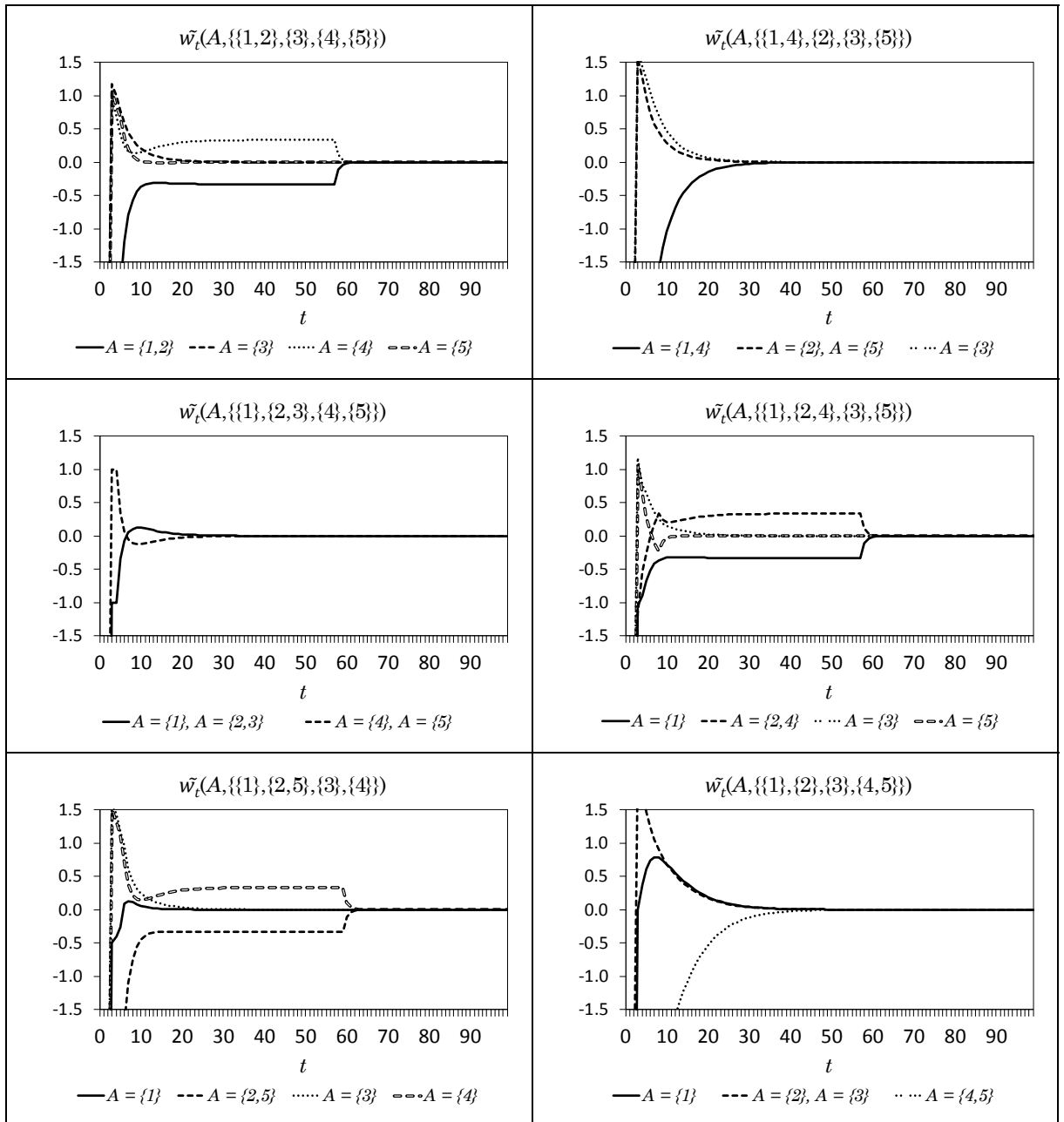with $w_0 \equiv \mathcal{S}(v)$ and $|\pi| = 5$.

Figure 3: $\widetilde{w_t}(A, \pi) \equiv w_t(A, \pi) - \Phi(v)(A, \pi)$
with $w_0 \equiv \mathcal{S}(v)$ and $|\pi| = 4$.