# Coalition formation problems with externalities

María Haydée Fonseca-Mairena* Matteo Triossi†

**Abstract**

We study coalition formation problems with general externalities. We prove that if expectations are not prudent a stable coalitions structure can fail to exist. Under prudent expectations a stable coalition structure exists if the set of admissible coalitions is single-lapping. This assumption also guarantees the existence of a stable and efficient coalition structure. However, under this assumption, the stable set is not a singleton, and no stable and efficient strategy-proof revelation mechanism exists, differently from the case in which agents care only about the coalition they belong to. However, the stable correspondence is implementable in Nash equilibrium.

Keywords: Coalition formation problems, Externalities, Stability, Efficiency.

Economic Literature Classification Numbers: C71, C72, C78, D62, D78.

*Corresponding author. Department of Economics and Management, Universidad Católica del Maule. San Miguel 3605, Talca, Chile. E-mail: mfonseca@ucm.cl.

†Department of Management, Ca' Foscari University of Venice. Fondamenta San Giobbe, Cannaregio 873, 30121 Venice, Italy. E-mail: matteo.triossi@unive.it.

# 1 Introduction

We study the existence of stable coalition structures in coalition formation problems with general externalities, in which agents also care about the coalitions they do not belong to. Partner dance competitions and team sports competitions, in which each member of the couple/team cares about how the other couples/teams are formed because it will affect their rank in the competitions, are appropriate examples. Other examples are cartel formation models in which the profits of the members of a cartel also depend on fringe firms and on the members of the other cartels, and international trade models in which the gain from fiscal coordination between a group of countries depends also on the behavior of the countries outside the alliance.

Stability is a focal concept in coalition formation theory. A coalition structure, which is a partition of the sets of all agents, is stable if there exists no admissible coalition whose agents prefer to group together rather than join the coalition assigned to them by the coalition structure (see Pápai, 2004). Defining stability in markets with general externalities involves modeling the agents' expectations about the behavior of the other agents outside their coalition (see Bando, Kawasaki, and Muto, 2016). We consider a concept of stability based on prudent expectations and assume that a coalition blocks a given coalition structure if its members prefer to stay together rather than accepting the coalition assigned by the initial coalition structure, independently on how the agents outside the coalition regroup (see Sasaki and Toda, 1996).

Our analysis relies on a restriction on the set of admissible coalitions called the *single-lapping* property (see Pápai, 2004). It is essentially the unique restriction that guarantees the strategy-proofness of stable revelation mechanisms in coalition formation problems without externalities (see Pápai, 2004 and Rodríguez-Álvarez, 2009). Restricting the set of admissible coalitions is equivalent to restricting preferences, for example, by making coalitions that are not permissible not individually rational. However, from a market design perspective, restricting the set of admissible preferences is more problematic than restricting the set of admissible coalitions, which can be implemented in practice.

We tackle three complementary issues: the existence of stable coalition structures, their efficiency, and their strategy-proofness. We start by proving that, if the expectations of the agents are not prudent the stable set may be empty. Then, we prove that under prudent expectation the set of stable coalition structures is nonempty if the set of admissible coalitions is single-lapping. We show, by means of an example, that the stable coalition structures set is not, in general, Pareto efficient. However, the set of Pareto efficient stable coalition structures is nonempty if the set of admissible coalitions is single-lapping. A single-lapping set of admissible coalitions does not guarantee the existence of a strategy-proof stable revelation mechanism nor that the set of stable coalition structures is a singleton, differently from the model in which agents care only about the coalition they belong to (Pápai, 2004 and Rodríguez-Álvarez, 2009). However, if we restrict our attention to strict preferences only, the stable correspondence is implementable in Nash equilibrium.

The paper is organized as follows. In Section 2 we introduce the model. In Section 3 we present the results. In Section 4 we conclude.

## 2 The model

A *coalition formation problem with externalities* is a tuple $\mathcal{A} = (N, \Pi^*, R)$, in which $N = \{1, \ldots, n\}$ is the finite set of agents, $\Pi^* \subseteq 2^N \backslash \{\varnothing\}$ is the set of admissible coalitions, and $R = (R_i)_{i \in N}$ is the profile of agents' preferences.

**Example 1** *The model is apt to represent well-known problems such as:*

*(i) Roommate problems:* $\Pi^* \subseteq \{T : T \subseteq N, |T| \in \{1, 2\}\}$.

*(ii) Marriage markets: let* $W$ *and* $M$ *be two finite and nonempty sets of agents such that* $W \cup M = N$, $W \cap M = \varnothing$, *and* $\Pi^* \subseteq \{T : (i, j) \in T, (i, j) \in (W \cup \{j\}) \times (M \cup \{i\})\}$.

A coalition structure is a partition of $N$, $\sigma = \{T_1, \ldots, T_k\} \subseteq \Pi^*$ with $n \geqslant k \geqslant 1$. Every agent $i \in N$ has complete and transitive preferences, $R_i$, over $\Sigma$, the set of all coalition structures. We assume $e = \{\{i\}\}_{i \in N} \in \Sigma$. Let $P_i$ and $I_i$ denote the asymmetric and symmetric parts of $R_i$, respectively. Let $\mathcal{R}_i$ be the set of preferences of agent $i$. Let

$\mathcal{R} = \prod_{i \in N} \mathcal{R}_i$ be the set of preference profiles. Let $\mathcal{P}$ be the set of strict preference profiles. Differently from classical models of coalition formation, the preferences of each agent depend also on the coalitions she does not belong to. All along the paper, we only require that no agent is indifferent between two coalition structures that assign her to different coalitions.

**Assumption 1** *For all $i \in N$, for all $R_i \in \mathcal{R}_i$, and for all $\sigma, \sigma' \in \Sigma$, $\sigma I_i \sigma' \implies \sigma(i) = \sigma'(i)$.*

Let $\Sigma(T) = \{\sigma \in \Sigma : T \in \sigma\}$ be the set of coalition structures in which the members of coalition $T$ stay together. Let $\varphi_i(T) \subseteq \Sigma(T)$ be the set of coalition structures agent $i$ expects if coalition $T$ deviates. When coalition $T$ blocks a coalition structure, the agents outside the coalition may react by regrouping in several manners. Thus, the agents must consider their beliefs regarding the possible reactions. Formally, given $R \in \mathcal{R}$, coalition $T \subseteq \Pi^*$ *blocks* $\sigma$ if for all $i \in T$, $\sigma' R_i \sigma$ for all $\sigma' \in \varphi_i(T)$, and there exists some $j \in T$ such that $\sigma' P_j \sigma$ for all $\sigma' \in \varphi_j(T)$.

If $\varphi_i(T) = \Sigma(T)$ for all $i \in N$, we say that agents have prudent expectations. This means that they believe that the agents outside $T$ can regroup in any possible way. Thus, they deviate if and only if the worst possible payoff from the deviation makes no agent in $T$ worse off and makes at least one agent in $T$ strictly better off with respect to the proposed coalition structure.

A coalition structure $\sigma \in \Sigma$ is *stable* under $R$ if no coalition blocks it. Let $S(R)$ be the set of stable coalition structures at $R$, which defines a correspondence $S : \mathcal{R} \rightrightarrows \Sigma$. A coalition structure $\sigma \in \Sigma$ is *efficient* under $R$ if there is no other coalition structure $\sigma' \in \Sigma$ such that $\sigma' R_i \sigma$ for all $i \in N$ and $\sigma' P_j \sigma$ for some $j \in N$. Let $PO(R)$ be the set of efficient coalition structures under $R$.

A revelation mechanism $\Gamma : \mathcal{R} \to \Sigma$ is *stable* if $\Gamma(R) \in S(R)$ for all $R \in \mathcal{R}$ and is *efficient* if $\Gamma(R) \in PO(R)$ for all $R \in \mathcal{R}$. A revelation mechanism is *strategy-proof* if, for all $i \in N$, $\Gamma(R_i, R_{-i}) R_i \Gamma(R'_i, R_{-i})$ for all $(R_i, R_{-i})$, $(R'_i, R_{-i}) \in \mathcal{R}$. Let $\mathcal{D} \subseteq \mathcal{R}$ and let $\Omega : \mathcal{D} \rightrightarrows \Sigma$ be a correspondence. A mechanism is a pair $(M, g)$ where $M = \prod_{i \in N} M_i$, $M_i$ is the strategy space of agent $i \in N$ and $g : M \to \Sigma$ is the outcome function. For all $R \in \mathcal{D}$,

$(M, g)$ induces a strategic form game, $(N, R, M, g)$. We denote by $NE(R, M, g)$ the set of pure strategy Nash equilibria of game $(N, R, M, g)$. Mechanism $(M, g)$ implements $\Omega$ in Nash equilibrium if $g(NE(R, M, g)) = \Omega(R)$ for all $R \in \mathcal{D}$.

The set of admissible coalitions is single-lapping if it satisfies two properties. First, any two coalitions share at most one agent. Second, any sequence of overlapping coalitions shares exactly the same agent.

**Definition 1** *The set of admissible coalitions* $\Pi^*$ *is single-lapping if*

*(i) for all* $T, T' \in \Pi^*$ *such that* $T \neq T'$, $|T \cap T'| \leqslant 1$, *and*

*(ii) for all* $\{T_1, \ldots, T_m\} \subseteq \Pi^*$ *such that* $m \geqslant 3$ *and for all* $l = 1, \ldots, m$, $|T_l \cap T_{l+1}| \geqslant 1$, *where we let* $T_{m+1} := T_1$, *there exists* $i \in N$ *such that for all* $l = 1, \ldots, m$, $T_l \cap T_{l+1} = \{i\}$.

Theorem 1 in Pápai (2004) proves that, if agents only care about the coalition they belong to, a coalition formation model has a unique stable coalition structure if and only if the set of admissible coalitions is single-lapping.[1]

## 3 Results

We first present a necessary condition for the nonemptiness of the stable set. If agents are not prudent, the set of stable coalition structures is empty for at least one preference profile, which generalizes Proposition 3.1 in Sasaki and Toda (1996) beyond the marriage market case.

**Proposition 1** *Let* $n \geqslant 3$. *If* $\varphi_i(T) \neq \Sigma(T)$ *for some* $i \in T$ *and some* $T \in \Pi^*$, *there exists a preference profile* $R \in \mathcal{R}$ *such that* $S(R) = \varnothing$.

**Proof.** Assume there exist $T^* \in \Pi^*$ and $i^* \in T^*$ such that $\varphi_{i^*}(T^*) \neq \Sigma(T^*)$. Let $\sigma^* \in \Sigma(T^*) \backslash \varphi_{i^*}(T^*)$. Let $R \in \mathcal{R}$ such that, for all $T \in \sigma^*$, for all $i \in T$, $\widetilde{\sigma} P_i \sigma^* P_i \widetilde{\sigma}'$ for

[1] In roommate problems, the single-lapping condition strengthens the no-odd-ring condition by Chung (2000). See Pápai (2004) for a full discussion of this issue and Contreras and Torres-Martínez (2021) for an application to roommate problems with externalities and prudent agents.

all $\widetilde{\sigma} \in \varphi_i(T) \backslash \{\sigma^*\}$ and $\widetilde{\sigma}' \notin \varphi_i(T)$. Let $\sigma \neq \sigma^*$ and let $T \in \sigma^* \backslash \sigma$. Then $T$ blocks $\sigma$. Moreover, $T^*$ blocks $\sigma^*$ because $\sigma^* \notin \varphi_{i*}(T^*)$, which implies that $S(R) = \varnothing$. ■

The result is independent on any assumption on the structure of the set of admissible coalitions. From now on, we assume agents are prudent. In the marriage model, stable coalition structures always exist under prudent expectations (Sasaki and Toda, 1996). However, in general models of coalition formation, stable coalition structures may fail to exist as shown in Example 2.

**Example 2** *Let* $N = \{1, 2, 3, 4\}$ *and*

$$\Pi^* = \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\},\{1\},\{2\},\{3\},\{4\}\}.$$

*Consider the following coalition structures:*

$$e = \sigma_0 = \{\{1\},\{2\},\{3\},\{4\}\}, \qquad \sigma_1 = \{\{1,2\},\{3\},\{4\}\},$$

$$\sigma_2 = \{\{1,2\},\{3,4\}\}, \qquad \sigma_3 = \{\{1\},\{2\},\{3,4\}\},$$

$$\sigma_4 = \{\{1,3\},\{2,4\}\}, \qquad \sigma_5 = \{\{1,4\},\{2,3\}\}.$$

*The preferences are:*

$$\sigma_3 P_1 \sigma_2 P_1 \sigma_1 P_1 \sigma_4 P_1 \sigma_5 P_1 \sigma_0, \qquad \sigma_5 P_2 \sigma_2 P_2 \sigma_1 P_2 \sigma_4 P_2 \sigma_3 P_2 \sigma_0,$$

$$\sigma_4 P_3 \sigma_5 P_3 \sigma_2 P_3 \sigma_3 P_3 \sigma_1 P_3 \sigma_0, \qquad \sigma_5 P_4 \sigma_4 P_4 \sigma_2 P_4 \sigma_3 P_4 \sigma_1 P_4 \sigma_0.$$

*We have* $S(R) = \varnothing$ *because:*

*(i)* $\sigma_i$ *is blocked by* $\{2,3\}$*, for* $i \in \{0,1,2,3\}$,

*(ii)* $\sigma_4$ *is blocked by* $\{1,2\}$,

*(iii)* $\sigma_5$ *is blocked by* $\{1,3\}$.

In Example 2, $\Pi^*$ is not single-lapping. Instead, if the set of admissible coalitions is single-lapping, there exists a stable coalition structure.

**Theorem 1** *Let $\Pi^*$ be single-lapping. Then $S(R) \neq \emptyset$ for all $R \in \mathcal{R}$.*

**Proof.** Let $R \in \mathcal{R}$. We modify $R$ into a preference profile without externalities. For all $\sigma \in \Sigma$, let $\sigma^{w,R_i} \in \Sigma$ be the worst coalition structure according to $R_i$ in which agent $i$ keeps the same coalition as in $\sigma$: $\sigma^{w,R_i}(i) = \sigma(i)$ and $\sigma' R_i \sigma^{w,R_i}$ for all $\sigma'(i) = \sigma(i)$: $\sigma^{w,R_i}$. Let $R^* \in \mathcal{R}$ such that:

(a) for all $\sigma, \tau \in \Sigma$ such that $\sigma(i) = \tau(i)$, $\sigma I_i^* \tau$ ;

(b) for all $\sigma, \tau \in \Sigma$ such that $\sigma(i) \neq \tau(i)$, $\sigma P_i^* \tau$ if $\sigma^{w,R_i} P_i \tau^{w,R_i}$.

From Theorem 1 in Pápai (2004), $S(R^*) \neq \emptyset$. We prove by contradiction that $S(R^*) \subseteq S(R)$. Assume $\sigma \in S(R^*)$ but $\sigma \notin S(R)$. Then there exists $T \in \Pi^*$ such that $\sigma' P_i \sigma$ for all $i \in T$ and all $\sigma' \in \Sigma(T)$. In particular $\sigma'^{w,R_i} P_i \sigma^{w,R_i}$ for all $i \in T$ then $T$ blocks $\sigma$ under $R^*$ which yields a contradiction. ■

Differently from what happens in markets without externalities (Theorem 1 in Pápai, 2004), a single-lapping set of admissible coalitions does not guarantee that the set of stable coalition structures is a singleton. Moreover, as shown in the following example, a stable coalition structure may not be efficient.

**Example 3** *Let $N = \{1, 2, 3, 4\}$. Consider the set of admissible coalitions*

$$\Pi^* = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1,3\}, \{1,4\}, \{2,4\}\}.$$

*Notice that $\Pi^*$ is single-lapping. Let $\sigma_1 = \{\{1,3\}, \{2,4\}\}$, $\sigma_2 = \{\{1,3\}, \{2\}, \{4\}\}$, $\sigma_3 = \{\{1,4\}, \{2\}, \{3\}\}$, $\sigma_4 = \{\{2,4\}, \{1\}, \{3\}\}$, $\sigma_5 = e = \{\{1\}, \{2\}, \{3\}, \{4\}\}$. Consider the following preferences.*

$$\sigma_1 P_1 \sigma_4 P_1 \sigma_3 P_1 \sigma_2 P_1 \sigma_5, \qquad \sigma_3 P_2 \sigma_2 P_2 \sigma_1 P_2 \sigma_4 P_2 \sigma_5,$$

$$\sigma_1 P_3 \sigma_3 P_3 \sigma_4 P_3 \sigma_5 P_3 \sigma_2, \qquad \sigma_1 P_4 \sigma_2 P_4 \sigma_3 P_4 \sigma_4 P_4 \sigma_5.$$

*The associated preferences without externalities defined in the proof of Theorem 1 are:*

$$\sigma_3 P_1^* \sigma_1 I_1^* \sigma_2 P_1^* \sigma_4 I_1^* \sigma_5, \qquad \sigma_1 I_2^* \sigma_4 P_2^* \sigma_3 I_2^* \sigma_2 I_2^* \sigma_5,$$

$$\sigma_3 I_3^* \sigma_4 I_3^* \sigma_5 P_3^* \sigma_1 I_3^* \sigma_2, \qquad \sigma_3 P_4^* \sigma_1 I_4^* \sigma_4 P_4^* \sigma_2 I_4^* \sigma_5.$$

*We have* $S(R^*) = \{\sigma_3\} \subset S(R) = \{\sigma_1, \sigma_3, \sigma_4\}$.[2]

In Example 3, although $\sigma_4$ is not efficient under $R$, there are coalition structures that are stable and efficient, $\sigma_1$ and $\sigma_3$. Indeed, the next claim shows that always exists a stable and efficient coalition structure in any coalition formation problems with externalities.

**Theorem 2** *For any coalition formation problem with externalities* $\mathcal{A}(N, \Pi^*, R)$ *such that* $\Pi^*$ *is single-lapping, the set* $S(R) \cap PO(R) \neq \varnothing$.

**Proof.** By Theorem 1, $S(R) \neq \varnothing$. Assume $\sigma \in S(R)$ such that $\sigma \notin PO(R)$. Then, there exists $\sigma' \in \Sigma$ such that $\sigma' P_i \sigma$ for all $i \in N$. Next, we prove $\sigma' \in S(R)$. By contradiction assume $\sigma' \notin S(R)$. Then, there exists $T \in \Pi^*$ such that $\tau P_i \sigma'$ for all $i \in T$ and all $\tau \in \Sigma(T)$. By transitivity of $R$ we have $\tau P_i \sigma$ for all $i \in T$ and all $\tau \in \Sigma(T)$, then $\sigma \notin S(R)$ which yields a contradiction. ■

Next, we study incentives and implementation in coalition formations problems with externalities and prove that no stable, nor stable and efficient revelation mechanism is strategy-proof.

**Proposition 2** *Let* $|N| \geqslant 4$. *Then, there exists a single-lapping coalition formation problem such that*

*(i) no stable revelation mechanism is strategy-proof,*

*(ii) no stable and efficient revelation mechanism is strategy-proof.*

**Proof.** Consider the coalition formation problem of Example 3. We have $S(R) = \{\sigma_1, \sigma_3, \sigma_4\}$ and $S(R) \cap PO(R) = \{\sigma_1, \sigma_3\}$. Let $\Gamma : \mathcal{R} \to \Sigma$ be a revelation mechanism. Consider the following cases.

[2] Under $R^*$ all coalition structures except $\sigma_3$ are blocked by {1,4}. Under $R$, $\sigma_2$ is blocked by agent 3 and $\sigma_5$ is blocked by $\{1, 4\}$ and $\{2, 4\}$.

(a) $\Gamma(R) = \{\sigma_3\}$. Let $\widetilde{R} = \left(\widetilde{R}_1, R_{-1}\right)$ in which $\widetilde{R}_1 = \sigma_1 \widetilde{P}_1 \sigma_4 \widetilde{P}_1 \sigma_5 \widetilde{P}_1 \ldots$. Then $S\left(\widetilde{R}\right) = \{\sigma_1, \sigma_4\}$. Since $\sigma' P_1 \sigma_3$ for all $\sigma' \in S\left(\widetilde{R}\right)$, agent 1 has incentives to misrepresent her preferences.

(b) $\Gamma(R) \in \{\sigma_1, \sigma_4\}$. Let $\widetilde{R} = \left(\widetilde{R}_2, R_{-2}\right)$ in which $\widetilde{R}_2 = \sigma_2 \widetilde{P}_2 \sigma_3 \widetilde{P}_2 \sigma_5 \widetilde{P}_2 \ldots$. Then $S\left(\widetilde{R}\right) = \{\sigma_3\}$. Since $\sigma_3 P_2 \sigma'$ for all $\sigma' \in \{\sigma_1, \sigma_4\}$, agent 2 has incentives to misrepresent her preferences.

If $\Gamma$ is a stable revelation mechanism, then $\Gamma(R) \in S(R)$, thus it cannot be strategy-proof. If $\Gamma$ is a stable and efficient revelation mechanism, then $\Gamma(R) \in S(R) \cap PO(R)$, thus it cannot be strategy-proof, which completes the proof of the claim for $N = 4$. The argument is easily generalized to the case $N > 4$. ■

The result holds even if we restrict our attention to strict preferences. However, if the preferences are strict and the set of admissible coalitions is single-lapping the stable set correspondence is implementable in Nash equilibrium.

**Proposition 3** *Let $\Pi^*$ be single-lapping. Then the restriction of the stable correspondence to the domain of strict preference profiles, $S_{|\mathcal{P}}$ is implementable in Nash equilibrium.*

The proof follows the same argument of the proof of Theorem 1 in Fonseca-Mairena and Triossi (2019) and is thus omitted.

# 4 Conclusions

In coalition formation problems with general externalities, assuming that agents have prudent expectations is a necessary condition for the nonemptiness of the stable set. If agents have prudent expectations and the set of admissible coalitions is single-lapping

(i) there exists a stable coalition structure,

(ii) the set of stable coalition structures is not a singleton nor efficient, in general, differently from what happens in markets without externalities,

(iii) a stable and efficient coalition structure always exists,

(iv) no stable, nor stable and efficient revelation mechanism is strategy-proof,

(v) the stable correspondence is implementable in Nash equilibrium under strict preferences.

### Acknowledgements

Triossi acknowledges financial support from Ca' Foscari University of Venice under project MAN.INS_TRIOSSI and from Ministerio Economía y Competitividad under projects MICIN Grants AEI/10.13039/501100011033 and AEI PID2020-118022GB-I0.

## References

[1] Bando, K., Kawasaki, R., Muto, S., 2016. Two-sided matching with externalities: A survey. Journal of the Operations Research Society of Japan 59, 35-71. https://doi.org/10.15807/jorsj.59.35

[2] Chung, K.S., 2000. On the Existence of Stable Roommate Matchings. Games and Economic Behavior 33, 206-230. https://doi.org/10.1006/game.1999.0779

[3] Contreras, J.L., Torres-Martínez, J.P., 2021. The roommate problem with externalities. International Journal of Game Theory 50, 149–165. https://doi.org/10.1007/s00182-020-00743-z

[4] Fonseca-Mairena, M.H., Triossi, M., 2019. Incentives and implementation in marriage markets with externalities. Economics Letters 185, 108688. https://doi.org/10.1016/j.econlet.2019.108688

[5] Pápai, S., 2004. Unique stability in simple coalition formation games. Games and Economic Behavior 48, 337-354. https://doi.org/10.1016/j.geb.2003.09.004

[6] Rodríguez-Álvarez, C., 2009. Strategy-proof coalition formation. Int. J. Game Theory 38, 431-452. https://doi.org/10.1007/s00182-009-0162-9

[7] Sasaki, H., Toda, M., 1996. Two-sided matching problems with externalities. Journal of Economic Theory 70, 93-108. https://doi.org/10.1006/jeth.1996.0077