Estabilidad e Incentivos en *Matching Markets* con Externalidades

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN ECONOMÍA

Alumno: Eduardo Duque Rosas
Profesor Guía: Juan Pablo Torres-Martínez

Santiago, Junio 2022

Estabilidad e Incentivos en *Matching Markets* con Externalidades*

Eduardo Duque Rosas**

Departamento de Economía
Universidad de Chile

Junio, 2022

1. Introducción

A causa de la gran importancia de la teoría de matching para emparejar individuos o distribuir objetos en ausencia de transferencias, preguntarse acerca de la incorporación de externalidades resulta igualmente relevante. Estudiantes que consideran las características de los alumnos en su establecimiento educacional o individuos que toman decisiones laborales basados en los atributos de sus colegas, son algunos ejemplos que motivan lo valioso de incorporar este análisis en la compresión de los matching y sus propiedades.

La inclusión de externalidades en el estudio de la teoría de matching requiere de dos elementos: (i) preferencias individuales que dependan de la distribución completa de los agentes en parejas, o de la distribución de objetos entre individuos; y (ii) creencias sobre las posibles reacciones del resto ante el desvío de un grupo para seguir un acuerdo particular.

Lamentablemente, sin hacer mayores supuestos acerca de la estructura de las preferencias o de las creencias, en los modelos que incluyan externalidades podrían no existir matchings estables o en el *core*. Por ejemplo, para el caso de *marriage markets*, Sasaki y Toda (1996) y Mumcu y Saglam (2010) muestran que no siempre existe un matching estable a desvíos de coaliciones, incluso en escenarios en los cuales los individuos son extremamente prudentes al momento de evaluar un desvío. Contreras y Torres-Martínez (2021) muestran que lo mismo ocurre para el caso de *roommate problems*. En el modelo habitacional de Shapley-Scarf, Mumcu y Saglam (2007), Aziz y Lee (2018), Graziano, Meo,

*Agradezco el constante apoyo de mi Profesor guía, Juan Pablo Torres-Martínez, por haberme permitido adentrarme un poco en el estudio de la teoría económica. También agradezco los valiosos comentarios de Jorge Arenas, Emilio Guamán y Emma Moreno-García. Los errores y omisiones son de mi única responsabilidad.

**eduquer@fen.uchile.cl

y Yannelis (2020), Klaus (2021), Klaus y Meo (2021) y Hong y Park (2022) muestran que el *core* puede ser vacío.

Es importante hacer notar que la ausencia de matchings estables tiende a disiparse a medida que la población crece. Efectivamente, Braitt y Torres-Martínez (2021) muestran que si las preferencias son generadas aleatoriamente y las externalidades arbitrarias, entonces la probabilidad de encontrar un matching estable para el caso de *marriage markets* como de *roommate problems*, converge a uno a medida que el tamaño de la población aumenta. En la misma linea de preferencias aleatorias y en un marco de un modelo general, Piazza y Torres-Martínez (2021) muestran que la probabilidad de tener un matching coalicionalmente estable está determinada por tres factores: prudencia de las coaliciones, conectividad social de quienes pueden reaccionar a un desvío, e incidencia de las externalidades en las preferencias. En este contexto, asumiendo que los individuos tienen una capacidad limitada para coordinares en grupos grandes, ellos muestran que la ausencia de matchings coalicionalmente estables tiende a disiparse con el crecimiento de la población, incluso en escenarios en que la prudencia de los individuos al momentos de evaluar un desvío cae a medida que la población aumenta.

Sin embargo, cuando el tamaño de la población está fijo, pueden aparecer problemas relevantes al momento de buscar una solución estable. Una forma de hacer frente estos problemas es dotar de cierta estructura a las preferencias de los agentes. En esta dirección, se han utilizado las llamadas *preferencias egocéntricas*. Con este tipo de preferencias, a los agentes les importa en primera instancia su pareja o el objeto que reciben y dejan en segundo plano la distribución del resto de los individuos u objetos. En general, el uso de preferencias egocéntricas ha traído consigo buenos resultados en la obtención de matchings estables. Esto se debe en gran medida a la poca relevancia que tienen las externalidades en las preferencias de los agentes cuando los individuos evalúan un desvío. Trabajos en los cuales se hace uso de preferencias próximas a preferencias egocéntricas son Li (1993) y Sasaki y Toda (1996) para el caso de problemas de matching bilateral. Para el caso de *Shapley-Scarf housing markets* ver Aziz y Lee (2018), Graziano, Meo, y Yannelis (2020), Klaus y Meo (2021), Klaus (2021) y Hong y Park (2022).[1]

A pesar de las semejanzas aparentes entre un escenario sin externalidades y un escenario en donde hay preferencias egocéntricas, puede que bajo estas últimas circunstancias existan problemas relacionados a los incentivos de los agentes al momento de reportar información sobre sus preferencias. En un modelo sin externalidades, la única forma que una persona tiene de mejorar su situación es cambiar de pareja. Cuando se incorporan externalidades al modelo, un individuo podría tener dos formas de mejorar su situación: cambiar de pareja o conseguir que otros cambien su situación. Por esta razón, podría ocurrir que un mecanismo que en la ausencia de externalidades es *strategy-proof* para un lado del mercado, pueda ser adaptado a un contexto de preferencias egocéntricas, pero deje de ser *strategy-proof* para ese lado del mercado por la presencia de factores externos.

[1]Por ejemplo, en el contexto del modelo habitacional del Shapley y Scarf (1974), Klaus y Meo (2021) presentan preferencias en donde a los agentes les importa en primer lugar quién se quedó con su objeto y en segundo lugar qué objeto obtendrán ellos.

De forma más concreta, en la ausencia de externalidades, el algoritmo de aceptación diferida propuesto por Gale y Shapley (1962) es *strategy-proof* para el lado del mercado que hace las propuestas (vea Dubins y Freedman (1981)). Sin embargo, aceptación diferida *no* es *non-bossy*, en el sentido que no es posible evitar que un individuo reporte preferencias falsas con el objetivo de cambiar la situación de otros individuos sin cambiar su propia pareja. Por esta razón, cuando las preferencias son egocéntricas—escenario en que el algoritmo de aceptación diferida puede ser adaptado de forma natural—los individuos podrían mejorar a través de cambiar las asignaciones de los otros.

Motivados por estas ideas, en la Sección 3 se estudiará el uso de preferencias egocéntricas en problemas de *marriage markets*. Como se mencionó, en un contexto de preferencias egocéntricas, puede que muchos resultados del caso sin externalidades sigan siendo válidos (vea Proposición 3.1). Sin embargo, como el principal mecanismo en el contexto de *marriage markets*—aceptación diferida—*no* es *non-bossy*, demostraremos que el mecanismo análogo para preferencias egocéntricas no es *strategy-proof* ni siquiera para aquellos agentes que hacen las propuestas (ver Teorema 1). Debido a lo anterior, nos preguntaremos si existe algún mecanismo, definido en un dominio que contenga los perfiles de preferencias egocéntricas, que sea *strategy-proof* al menos para los individuos de un lado del mercado. Inspirados por el resultado de Kojima (2010), que muestra que sin externalidades no existen mecanismos estables y *non-bossy*, llegaremos a una respuesta negativa a la pregunta anterior (ver Teorema 2). A pesar de aquello, mostraremos que existe un mecanismo, análogo a *top trading cycles*, que cumple con ser *strategy-proof*, Pareto eficiente e individualmente racional para un lado del mercado (ver Proposición 3.3).[2] Finalmente, inspirados por los resultados de Ergin (2002) para modelos sin externalidades, proponemos restricciones sobre el dominio de preferencias que aseguran que el mecanismo análogo a aceptación diferida no solo implementa soluciones estables, sino que es *strategy-proof*, *group strategy-proof* y Pareto eficiente para aquellos que hacen las propuestas (ver Teorema 3).

Otra forma de trabajar con externalidades es imponer estructura a las creencias de los individuos sobre las potenciales reacciones a desvíos. En la Sección 4 seguiremos este enfoque. Esto es, no impondremos restricciones sobre las preferencias, pero supondremos que los agentes tendrán en consideración *todos* los posibles escenarios al momento de evaluar un desvío. En este contexto, como demostraron Sasaki y Toda (1996), a partir de las preferencias sobre matchings es posible construir una preferencia sobre individuos. Aplicando el algoritmo de aceptación diferida de Gale y Shapley (1962) a estas últimas preferencias, siempre se obtiene un matching estable con externalidades (ver el Teorema 1 en Sasaki y Toda (1996)). Además, este mecanismo no es *strategy-proof* para ningún lado del mercado (ver Proposición 2 en Fonseca-Mairena y Triossi (2019)). Efectivamente, como los resultados que probamos en la Sección 3 no dependen de las creencias de los individuos, estos seguirán siendo válidos en el contexto propuesto por Sasaki y Toda (1996). Además, las restricciones sobre el dominio de preferencias que aseguraban buenas propiedades para la adaptación de aceptación diferida al contexto de preferencias egocéntricas dejan de cumplir ese rol para el mecanismo propuesto por Sasaki y Toda (1996) (ver Teorema 4).

[2] El mecanismo *top trading cycles* corresponde al desarrollado por Abdulkadiroğlu y Sönmez (2003).

En la Sección 5 proponemos restricciones a las preferencias que aseguran la existencia y unicidad de soluciones en el *core*. Esencialmente, asumiremos que es posible particionar la población de tal forma que todos los miembros de un grupo coincidan en posicionar los mismos matchings en el top de sus preferencias (ver Definiciones 5.1 y 5.2). Este tipo de estructuras nos asegurarán que el *core* tiene un único elemento (ver Teorema 5), aunque deja de coincidir con el conjunto de emparejamientos estables (ver Teorema 6). Así, se pierde el resultado que Roth y Sotomayor (1990) probaron para modelos sin externalidades. Daremos condiciones suficientes para recuperar la igualdad entre el *core* y el conjunto de matchings estables en el marco de las externalidades estudiadas (ver Proposición 5.1).

Finalmente, en la Sección 6 propondremos una definición de *bargaining set* que es un refinamiento del concepto propuesto por Hafalir (2008). Esencialmente, permitimos que coaliciones que *contra-bloqueen* no necesariamente tengan creencias miopes sobre las potenciales reacciones de los otros individuos. En este contexto los resultados de existencia de *bargaining set* demostrados por Hafalir (2008) dejan de ser válidos. Esto es, el *bargaining set* puede ser vacío incluso cuando existe un matching que todos los individuos consideran como su peor alternativa (ver Proposición 6.1).

El resto del texto está organizado de la siguiente forma. En la Sección 2 se detallará el modelo, así como los conceptos utilizados. La Sección 3 tratará sobre el uso de referencias egocéntricas en el contexto de *marriage markets*. Se describirá de manera formal este tipo de preferencias, como también los mecanismos a analizar y sus propiedades. En la Sección 4 se asumirá que los individuos son extremamente prudentes y se caracterizarán las propiedades del mecanismo estable propuesto por Sasaki y Toda (1996). La Sección 5 se enfocará en el análisis de estabilidad bajo dos nuevos tipos de estructura de preferencias, las cuales nos asegurarán no solo la existencia sino la unicidad de soluciones en el *core*. En la Sección 6 se introducirá un nuevo concepto de *bargaining set* y se comparará con el propuesto por Hafalir (2008). Algunas de las demostraciones se dejarán en el Apéndice A.

2. Modelo

En un *marriage market problem*

$$\Lambda = \left\langle M, W, (R_i)_{i \in M \cup W}, \left(\Theta^{T,f}\right)_{T \subseteq M \cup W, f \in \mathcal{A}(T)} \right\rangle$$

existen dos conjuntos finitos de agentes M y W, donde pueden formar parejas entre miembros de distinto grupo o estar solos. Un matching es una función $\mu : M \cup W \to M \cup W$ en donde para todo $m \in M$ y todo $w \in W$ se tiene

- $\mu(m) \in W \cup \{m\}$.
- $\mu(w) \in M \cup \{w\}$.
- $\mu(m) = w$ si y solamente si $\mu(w) = m$.

Denotaremos por $\mathcal{A}(M, W)$ al conjunto de todos los matchings entre los miembros de M y W. Asumiremos que cada individuo $i \in M \cup W$ tiene preferencias R_i definidas sobre los elementos del conjunto $\mathcal{A}(M, W)$. Esto es, dados $\mu, \eta \in \mathcal{A}(M, W)$, $\mu R_i \eta$ significa que al agente i le gusta el matching μ tanto como η. Asumiremos que cada R_i pertenece a un conjunto de preferencias $\mathcal{R}_i$, las cuales son completas y transitivas. Para cada $R_i \in \mathcal{R}_i$, denotaremos por P_i e I_i a sus partes asimétrica y simétrica, respectivamente.

Una coalición es un subconjunto no vacío de $M \cup W$. Dada una coalición T, nos referiremos a cualquier vector $R_T = (R_i)_{i \in T}$ como un *perfil de preferencias de* T y denotaremos por $\mathcal{R}_T = \prod_{i \in T} \mathcal{R}_i$ al conjunto de todos los perfiles de preferencias de T. Del mismo modo, diremos que R es un *perfil de preferencias* si R es un perfil de preferencias de $M \cup W$. Llamaremos $\mathcal{R} = \prod_{i \in M \cup W} \mathcal{R}_i$ al conjunto de todos los perfiles de preferencias.

Dada una coalición T, denote por $\mathcal{A}(T)$ al conjunto de matchings en $\mathcal{A}(M \cap T, W \cap T)$. Dado $f \in \mathcal{A}(T)$, denotaremos por

$$\mathcal{A}^{T,f} = \{\mu \in \mathcal{A}(M, W) : \mu(i) = f(i), \forall i \in T\}$$

al conjunto de matchings en los cuales los miembros de T son distribuidos como en f.

Dada una coalición $T \subseteq M \cup W$ y un potencial acuerdo $f \in \mathcal{A}(T)$ entre sus miembros para desviar de $\mu \in \mathcal{A}(M, W)$, denotaremos por $\Theta^{T,f}(\mu) \subseteq \mathcal{A}^{T,f}$ a la colección de matchings que los individuos en T piensan que pueden ser alcanzados como resultado de la reacción de los otros agentes a sus acciones. Por lo tanto, la correspondencia no vacía $\Theta^{T,f} : \mathcal{A}(M, W) \twoheadrightarrow \mathcal{A}^{T,f}$ determina las creencias de los agentes de T sobre las posibles reacciones del resto cuando ellos desvían para seguir el acuerdo f.

Definición 2.1. Un matching μ es *bloqueado* por una coalición T cuando existe un acuerdo $f \in \mathcal{A}(T)$ tal que:

1. $\mu(i) \neq f(i)$ para algún $i \in T$.
2. $\eta R_i \mu$ para todo $\eta \in \Theta^{T,f}(\mu)$ y para todo $i \in T$.
3. Para todo $\eta \in \Theta^{T,f}(\mu)$ existe un $i \in T$ que $\eta P_i \mu$.

Un matching está en el Θ-*core* si no es bloqueado por ninguna coalición T. Dado un perfil de preferencias $R \in \mathcal{R}$, denotaremos por $\mathcal{C}(R)$ al conjunto de matchings en el $\Theta-$*core*.

Un elemento a destacar en la Definición 2.1, es el hecho de que si una coalición bloquea, entonces alguien debe cambiar de pareja. Esto evita que existan coaliciones que bloqueen manteniendo sus parejas y mejoren producto de la reacción del resto. Esto toma gran importancia a la hora de considerar externalidades, las cuales puede comprometer la existencia de un $\Theta-$core distinto de vacío, incluso en el contexto de preferencias egocéntricas.[3]

A lo largo del texto utilizaremos los siguientes conceptos:

- Dado $T \subseteq M \cup W$, un matching μ es *individualmente racional para T* si no es bloqueado por coaliciones compuestas de un único agente de T.
- Un matching es *individualmente racional* si es individualmente racional para $M \cup W$.
- Un matching es *estable* si es individualmente racional y no es bloqueado por ninguna coalición $T = \{m, w\}$ a través del acuerdo $f \in \mathcal{A}(T)$ caracterizado por $f(m) = w$. Dado $R \in \mathcal{R}$, denotaremos por $\mathcal{S}(R)$ al conjunto de matchings estables cuando el perfil de preferencias es R.
- Dado $T \subseteq M \cup W$, un matching μ es *Pareto eficiente para T* si no existe $\mu' \in \mathcal{A}(M, W)$ tal que:
    - $\mu' R_i \mu$ para todo $i \in T$.
    - $\mu' P_i \mu$ para un $i \in T$.
- Un matching es *Pareto eficiente* si es Pareto eficiente para $M \cup W$.
- Dado $T \subseteq M \cup W$, un matching μ es *débilmente Pareto eficiente para T* si no existe $\mu' \in \mathcal{A}(M, W)$ tal que $\mu' P_i \mu$ para todo $i \in T$.
- Un matching es *débilmente Pareto eficiente* si es débilmente Pareto eficiente para $M \cup W$.

Un *mecanismo* es una función $\varphi : \mathcal{R} \to \mathcal{A}(M, W)$ que selecciona un matching para cada perfil de preferencias. Dado un perfil de preferencias $R = (R_i)_{i \in M \cup W}$, denotaremos por $\varphi(R)(i)$ a la pareja del individuo i en el matching $\varphi(R)$. Además, dado $S \subseteq M \cup W$, denotaremos por $R_{-S} = (R_j)_{j \in M \cup W \setminus S}$ al perfil de preferencias de los individuos j que no están en S.

Dado un mecanismo $\varphi : \mathcal{R} \to \mathcal{A}(M, W)$ diremos que:

- Dado $T \subseteq M \cup W$, φ es *strategy-proof para T* si se cumple la siguiente propiedad:
  $$\varphi(R) R_i \varphi(\tilde{R}_i, R_{-i}), \qquad \forall R \in \mathcal{R}, \forall i \in T, \forall \tilde{R}_i \in \mathcal{R}_i.$$
  Note que, cuando φ es *strategy-proof* para T, es una estrategia débilmente dominante para cada $i \in T$ reportar sus verdaderas preferencias.
- φ es *strategy-proof* si φ es *strategy-proof* para $M \cup W$.
- Dado $T \subseteq M \cup W$, φ es *group strategy-proof para T* si para cualquier $R \in \mathcal{R}$ y cualquier $S \subseteq T$, no existe $\tilde{R}_S \in \mathcal{R}_S$ que cumpla con las siguientes propiedades:
  $$\begin{aligned} &\varphi(\tilde{R}_S, R_{-S}) R_i \varphi(R), && \forall i \in S \\ &\varphi(\tilde{R}_S, R_{-S}) P_i \varphi(R), && \text{para algún } i \in S. \end{aligned}$$

[3] En el contexto de *Shapley-Scarf housing market*, el Ejemplo 6 de Fonseca-Mairena y Triossi (2022) muestra un caso de *core* vacío.

Note que, cuando φ es *group strategy-proof* para T, es una estrategia débilmente dominante para toda coalición $S \subseteq T$ reportar sus verdaderas preferencias.

- φ es *group strategy-proof* si φ es *group strategy-proof* para $M \cup W$.
- Dado $T \subseteq M \cup W$, φ es *non-bossy para T* si se cumple la siguiente propiedad:

$$\varphi(\tilde{R}_i, R_{-i})(i) = \varphi(R)(i) \implies \varphi(\tilde{R}_i, R_{-i}) = \varphi(R), \qquad \forall R \in \mathcal{R}, \forall i \in T, \forall \tilde{R}_i \in \mathcal{R}_i.$$

Note que, cuando φ es *non-bossy* para T, ningún individuo $i \in T$ puede cambiar el resultado del mecanismo sin afectar su propia situación.

- φ es *non-bossy* si φ es *non-bossy* para $M \cup W$.
- Dado $T \subseteq M \cup W$, φ es *group non-bossy para T* si para toda coalición $S \subseteq T$ se cumple la siguiente propiedad:

$$\varphi(\tilde{R}_S, R_{-S})(i) = \varphi(R)(i) \quad \forall i \in S \implies \varphi(\tilde{R}_S, R_{-S}) = \varphi(R), \qquad \forall R \in \mathcal{R}, \forall \tilde{R}_S \in \mathcal{R}_S.$$

Note que, cuando φ es *group non-bossy* para T, ninguna coalición $S \subseteq T$ puede cambiar el resultado del mecanismo sin afectar su propia situación.

- φ es *group non-bossy* si φ es *group non-bossy* para $M \cup W$.

Las propiedades de racionalidad individual y Pareto eficiencia se extenderán de matchings a mecanismos de forma natural. Efectivamente, dado un mecanismo $\varphi : \mathcal{R} \to \mathcal{A}(M, W)$ diremos que:

- Dado $T \subseteq M \cup W$, φ es *individualmente racional para T* si, para cada perfil de preferencias $R \in \mathcal{R}$, el matching $\varphi(R)$ es individualmente racional para T.
- φ es *individualmente racional* si $\varphi(R)$ es individualmente racional para todo $R \in \mathcal{R}$.
- φ es *estable* si $\varphi(R)$ es estable para todo $R \in \mathcal{R}$.
- φ es *Θ-core selecting* si $\varphi(R)$ está en el Θ-core para todo $R \in \mathcal{R}$.
- Dado $T \subseteq M \cup W$, φ es *Pareto eficiente para T* si, para cada perfil de preferencias $R \in \mathcal{R}$, el matching $\varphi(R)$ es Pareto eficiente para T.
- φ es *Pareto eficiente* si $\varphi(R)$ es Pareto eficiente para todo $R \in \mathcal{R}$.
- Dado $T \subseteq M \cup W$, φ es *débilmente Pareto eficiente para T* si, para cada perfil de preferencias $R \in \mathcal{R}$, el matching $\varphi(R)$ es débilmente Pareto eficiente para T.
- φ es *débilmente Pareto eficiente* si $\varphi(R)$ es débilmente Pareto eficiente para todo $R \in \mathcal{R}$.

3. Preferencias Egocéntricas en *Marriage Market*

Sin hacer mayores supuestos acerca de las creencias de los agentes,[4] Sasaki y Toda (1996) muestran que no se puede asegurar la existencia de un matching estable. Una de las formas para intentar solucionar aquello, es realizar supuestos acerca de la estructura de las preferencias. En esta sección se estudiarán un tipo particular de preferencias, la cual llamaremos *preferencias egocéntricas*. Por preferencias egocéntricas entenderemos que a los individuos, en primera instancia les importará su pareja y en segundo orden la distribución de parejas del resto. Veremos que el uso de preferencias egocéntricas, por más "suaves" que sean las externalidades en estas preferencias, provocarán problemas en los incentivos de los agentes al momento de reportar sus preferencias.

3.1. Definiciones

En este apartado describiremos de manera formal las preferencias egocéntricas. Como se mencionó anteriormente, la estructura de las preferencias estudiadas en esta sección, reflejarán que en primera instancia a las personas les importará su pareja y luego la distribución de parejas del resto. Basados en Hong y Park (2022) damos la siguiente definición.

Definición 3.1. La preferencia R_i es *egocéntrica* si, dados $\mu, \eta \in \mathcal{A}(M, W)$, se cumple:

- $\mu I_i \eta \Rightarrow \mu(i) = \eta(i)$.
- $\mu P_i \eta$ y $\mu(i) \neq \eta(i)$ implica que $\mu' P_i \eta'$ para todo μ' y η' tales que $\mu'(i) = \mu(i)$ y $\eta'(i) = \eta(i)$.

Notar que si $\mu(i) \neq \eta(i)$ entonces se tendrá $\mu P_i \eta$ o $\eta P_i \mu$. Dada una coalición T, si R_i es egocéntrica para todo $i \in T$, entonces diremos que $R_T = (R_i)_{i \in T}$ es un *perfil de preferencias egocéntricas de* T. $\mathcal{R}_T^{ego}$ es el conjunto de perfiles de preferencias egocéntricas de T. Del mismo modo, diremos que R es un *perfil de preferencias egocéntricas* si R es un perfil de preferencias egocéntricas de $M \cup W$. Denotaremos por $\mathcal{R}^{ego}$ al conjunto de perfiles de preferencias egocéntricas.

Dado $m \in M$, toda preferencia egocéntrica R_m induce una relación de preferencias *estándar*. Esta nueva preferencia, la cual denotaremos por R_m^s, está definida sobre $W \cup \{m\}$ y es tal que, dados $i, j \in W \cup \{m\}$ con $i \neq j$, se cumple la siguiente propiedad

$$i \, R_m^s \, j \qquad \Longleftrightarrow \qquad \mu P_m \eta, \qquad \forall \mu, \eta \in \mathcal{A}(M, W): \ \mu(m) = i, \eta(m) = j$$

Note que, gracias a que R_m es egocéntrica, R_m^s es completa, transitiva y estricta. Dada una coalición T, si para todo $i \in T$, R_i^s es la preferencia estándar inducida de alguna preferencia egocéntrica, entonces diremos que $R_T^s = (R_i^s)_{i \in T}$ es un *perfil de preferencias estándar de* T. Denotaremos por $\mathcal{R}_T^s$ al conjunto de perfiles de preferencias estándar de T. R^s es un *perfil de preferencias estándar* si R^s es un perfil de preferencias estándar de $M \cup W$. Llamaremos $\mathcal{R}^s$ al conjunto de perfiles de preferencias estándar. Para cada $R_i^s \in \mathcal{R}_i^s$, denotaremos por P_i^s e I_i^s a sus partes asimétrica y simétrica, respectivamente.

[4] En este texto, por creencias o *beliefs*, no referiremos las creencias de los agentes de una coalición sobre las posibles reacciones del resto cuando ellos desvían para seguir el acuerdo particular.

3.2. Mecanismos

En esta subsección introduciremos los mecanismos que se estudiarán en los siguientes apartados. Debido a que el dominio de preferencias son sobre matchings y no individuos, los algoritmos tradicionales como aceptación diferida (AD) o top trading cycle (TTC) no son directamente aplicables en este contexto.[5]

Primero, describiremos una variación del mecanismo AD,[6] al cual llamaremos *algoritmo de aceptación diferida con preferencias egocéntricas* (AD^e).[7] A diferencia de AD, esta variación tiene como dominio a perfiles preferencias egocéntricas. A pesar de aquello, veremos que a la hora de ejecutar este mecanismo, solo se tomará en consideración la parte estándar de las preferencias egocéntricas.[8] Lo anterior implica que, para un perfil de preferencias egocéntricas R y el perfil de preferencias estándar inducido R^s, se tendrá

$$AD^e_M(R) = AD_M(R^s).$$

Esto es, las etapas del mecanismo $AD^e_M : \mathcal{R}^{ego} \to \mathcal{A}(M, W)$ son las siguientes.

- *Etapa 1-a:* Cada $m \in M$ le hace una propuesta al más preferido de los $w \in W \cup \{m\}$ según la preferencia R^s_m si hay alguno aceptable, donde R^s_m es la preferencia estándar inducida de $R_m \in \mathcal{R}^{ego}_m$.
- *Etapa 1-b:* Cada $w \in W$ acepta temporalmente la más atractiva de las propuestas que recibe según la preferencia R^s_w, rechazando las otras y todas aquellas que son inaceptables. R^s_w es la preferencia estándar inducida de $R_w \in \mathcal{R}^{ego}_w$.
- *Etapa k-a:* Cada $m \in M$ que fue rechazado en la etapa previa, le hace una propuesta a la alternativa aceptable más preferida entre aquellas a las cuales aún no ha hecho propuestas según la preferencia R^s_m. Si no quedan alternativas aceptables, no hace propuestas.
- *Etapa k-b:* Cada $w \in W$ mantiene la alternativa más atractiva entre las que escogió en la etapa previa y las propuestas que recibe en esta etapa, según la preferencia R^s_w.

El algoritmo termina cuando no se hacen más propuestas.

En la misma linea de antes, presentaremos una variante del mecanismo TTC en el contexto de preferencias egocéntricas.[9] A este mecanismo lo llamaremos *top trading cycles con preferencias*

[5] A lo largo del texto nos referimos a AD_M (AD_W) como el mecanismo de aceptación diferida cuando M (W) hace las propuestas. También nombraremos por TTC_M (TTC_W) al mecanismo top trading cycle cuando se implementan las cadenas en donde los miembros de M (W) apuntan los miembros de $W(M)$.

[6] Ver Gale y Shapley (1962) para más detalles de este algoritmo.

[7] De igual forma, nos referimos a AD^e_M (AD^e_W) como el mecanismo de aceptación diferida con preferencias egocéntricas cuando M (W) hace las propuestas.

[8] Un ejemplo es lo realizado por Hong y Park (2022) en el contexto de Shapley- Scarf housing market para el caso del algoritmo TTC y preferencias egocéntricas.

[9] Ver Abdulkadiroğlu y Sönmez (2003) para más detalles de este algoritmo.

egocéntricas (TTC^e).[10] Veremos que a la hora de ejecutar este mecanismo, solo se tomará en consideración la parte estándar de las preferencias egocéntricas. Lo anterior implica que, para un perfil de preferencias egocéntricas R y el perfil de preferencias estándar inducido R^s, se tendrá

$$TTC_M^e(R) = TTC_M(R^s).$$

Esto es, las etapas del mecanismo $TTC_M^e : \mathcal{R}^{ego} \to \mathcal{A}(M, W)$ son las siguientes.

- *Etapa 1-a:* Cada $m \in M$ apunta al miembro preferido en $W \cup \{m\}$ según la preferencia R_m^s, donde R_m^s es la preferencia estándar inducida de $R_m \in \mathcal{R}_m^{ego}$.
- *Etapa 1-b:* Cada $w \in W$ apunta al miembro preferido en $M \cup \{w\}$ según la preferencia R_w^s, donde R_w^s es la preferencia estándar inducida de $R_w \in \mathcal{R}_w^{ego}$.

Como hay un número finito de agentes en $M \cup W$, se forma al menos un *ciclo*.[11] Se puede notar que cada $m \in M$ pertenecer a lo más a un ciclo. Por lo cual cada $m \in M$ que pertenece a un ciclo, se empareja con el miembro de $W \cup \{m\}$ al cual él apuntó y ambos son removidos del mercado.

- *Etapa k-a:* Cada $m \in M$ que aún no tiene pareja apunta al miembro preferido en $W \cup \{m\}$ entre los que aún no han salido, según la preferencia R_m^s.
- *Etapa k-b:* Cada $w \in W$ que aún no tiene pareja apunta al miembro preferido en $M \cup \{w\}$ entre los que aún no han salido, según la preferencia R_w^s.

Como hay un número finito de agentes se forma al menos un ciclo. Los miembros de M y W que aun no ha salido pertenecen a lo más a un ciclo. Los miembros de $m \in M$ que están a un ciclo, se empareja con el miembro de $W \cup \{m\}$ que apuntó y son removidos del mercado.

El algoritmo termina cuando todos los miembros de M han sido removidos.

3.3. Resultados

En esta subsección revisaremos los principales resultados relacionados a la incorporación de externalidades a través de preferencias egocéntricas en los modelos de *marriage markets*.

Al estudiar externalidades a través de preferencias egocéntricas, existen ciertas propiedades que resultan ser equivalentes respecto del caso sin externalidades. Por Roth y Sotomayor (1990) sabemos que en ausencia de externalidades el conjunto de matchings estables coincide con el conjunto de matchings en el *core*. En esta línea, se presentan las siguientes propiedades.

[10] Nos referimos a TTC_M^e (TTC_W^e) como el mecanismo top trading cycle con preferencias egocéntricas cuando se implementan las cadenas en donde los miembros de M (W) apuntan los miembros de $W(M)$.

[11] Un ciclo es una lista ordenada de miembros de M y miembros de W $(m_1, w_1, m_2, i_2, \ldots, m_k, w_k)$, en donde m_1 apunta a w_1, ..., m_k apunta a w_k, y w_k apunta a m_1.

Proposición 3.1. *Si $R \in \mathcal{R}^{ego}$ y R^s es el perfil de preferencias estándar inducido de R, entonces se cumple que $\mathcal{C}(R) = \mathcal{S}(R) = \mathcal{S}(R^s) = \mathcal{C}(R^s)$. Además, el mecanismo AD^e_M es $\Theta-$core selecting.*

Demostración. Ver Apéndice A. $\square$

Los siguientes resultados darán evidencia sobre las diferencias entre un escenario con preferencias egocéntricas y un escenario sin externalidades. En esta línea una primera diferencia entre ambos contextos está relacionada con la estructura de matching estables. Por Knuth (1976) sabemos que en ausencia de externalidades, los matching estables tienen una estructura de reticulado. Esto último, nos asegura la existencia de un matching estable que es el mejor para todos los miembros de M y de un matching estable (no necesariamente diferente al anterior) que es el mejor para todos los miembros de W. La siguiente proposición muestra que, incluso en escenarios en presencia de preferencias egocéntricas, dicha estructura se pierde.

Proposición 3.2. *Existen marriage markets con preferencias egocéntricas en los cuales no hay un matching estable que sea el mejor para todos los miembros de M.*

Demostración. Asuma que $M = \{m_1, m_2, m_3\}$ y $W = \{w_1, w_2, w_3\}$. Sea $R \in \mathcal{R}^{ego}$ un perfil de preferencias que induce preferencias estándar R^s que cumplen con las siguientes propiedades:

m_1	m_2	m_3	w_1	w_2	w_3
w_1	w_2	m_3	m_2	m_1	w_3
w_2	w_1		m_1	m_2	
m_1	m_2		w_1	w_2	

Dado que la Proposición 3.1 nos asegura que $\mathcal{S}(R) = \mathcal{S}(R^s)$, aprendemos que los únicos matching estables bajo la preferencia R son μ y μ', los cuales se describen a continuación:

$$\mu = \begin{bmatrix} m_1 & m_2 & m_3 & w_3 \\ w_1 & w_2 & m_3 & w_3 \end{bmatrix}, \qquad \mu' = \begin{bmatrix} m_1 & m_2 & m_3 & w_3 \\ w_2 & w_1 & m_3 & w_3 \end{bmatrix}.$$

Sin comprometer la descripción de las preferencias estándar, podemos asumir que $\mu' P_{m_3} \mu$ y $\mu P_{w_3} \mu'$. En este caso se tiene que μ es preferido por m_1 y m_2, pero m_3 prefiere μ' por sobre μ. Por lo tanto no hay un matching estable que sea el mejor para todos los miembros de M.

Note que en este mismo ejemplo tenemos que μ' es preferido por w_1 y w_2, pero w_3 prefiere μ por sobre μ'. En ambos casos vemos que no hay un matching estable que sea preferido por todos los miembros de W. $\square$

El resultado anterior muestra como la inclusión de pequeñas externalidades puede comprometer la estructura de matchings estables.

En la ausencia de externalidades, el algoritmo de aceptación diferida cuando los miembros de M hacen las propuestas—el cual hemos denotado por AD_M—ha sido ampliamente estudiado. En lo relacionado a la compatibilidad de incentivos, sabemos por el Teorema 9 de Dubins y Freedman (1981) que el mecanismo AD_M es *strategy-proof* para M. Por otro lado, no solo sabemos que AD_M es un mecanismo estable (Gale y Shapley, 1962, Teorema 1), sino también que el resultado de aplicar AD_M es el mejor de los matchings estables para los miembros de M (Gale y Shapley, 1962, Teorema 2). Finalmente, sabemos que AD_M es débilmente Pareto eficiente para M. Esto es, no existe matching individualmente racional (estable o no) que sea preferido sobre el resultado de aplicar AD_M para todos los miembros de M (Roth, 1982, Teorema 2).

A pesar de las buenas propiedades que posee AD_M, este mecanismo *no* es *non-bossy* para M. Esto es, a través de la manipulación de preferencias, individuos de M pueden cambiar las parejas del resto sin modificar su propia pareja. En un contexto en el que a los agentes les importa la distribución de parejas de los demás, como el caso de preferencias egocéntricas, no ser *non-bossy* podría generar incentivos a reportar preferencias falsas para así poder originar cambios en las parejas del resto y con eso beneficiarse. Como se comentó anteriormente, AD_M^e corresponde a un algoritmo análogo a AD_M, pero cuyo dominio son los perfiles preferencias egocéntricas. Veremos que el hecho de que AD_M no sea *non-bossy*, tendrá efectos en la eficiencia como en la compatibilidad de incentivos del mecanismo AD_M^e. Esto motiva el siguiente resultado.

Teorema 1. *El mecanismo* $AD_M^e : \mathcal{R}^{ego} \to \mathcal{A}(M, W)$ *cumple las siguientes propiedades:*

- AD_M^e *no siempre induce el mejor de los matchings estables para los miembros de* M.
- AD_M^e *siempre induce el mejor de los matchings estables para los miembros de* M*, caso de que dicho matching exista.*
- AD_M^e *no es strategy-proof para* M.
- AD_M^e *no es débilmente Pareto eficiente para* M.

Demostración. El hecho de que dado $R \in \mathcal{R}^{ego}$, $AD_M^e(R)$ no siempre induce el mejor de los matchings estables para todos los miembros de M, es una consecuencia directa de la Proposición 3.2 ya que puede que dicho matching no exista. Sabemos que AD_M^e siempre selecciona matchings estables. Sin embargo, en el caso de la demostración de la Proposición 3.2 no hay matching estable que sea el mejor para todos los miembros de M. En ese caso el resultado de aplicar AD_M^e es μ donde vemos que m_3 prefiere a μ'.

Asuma que μ es el mejor de los matchings estables para todos los miembros de M y que $AD_M^e(R) = \mu'$. Si $\mu \neq \mu'$, tenemos que $T = \{i \in M : \mu(i) \neq \mu'(i)\} \neq \emptyset$. Como para todo $i \in T$ se tiene que $\mu(i) \neq \mu'(i)$ y que $\mu P_i \mu'$, implica que $\mu(i) P_i^s \mu'(i)$, donde P_i^s es la preferencia estándar inducida del individuo i a partir de su preferencias egocéntrica. Como $\mu(i) I_i^s \mu'(i)$ para todo $i \in M \setminus T$ y que $AD_M^e(R) = AD_M(R^s)$, tenemos que sin externalidades AD_M no induce el mejor de los matchings estables para M, lo cual es una contradicción con el Teorema 2 de Gale y Shapley (1962).

El resto del Teorema se demostrará de la siguiente manera. Asuma que $M = \{m_1, m_2, m_3\}$ y $W = \{w_1, w_2\}$. Sea $R \in \mathcal{R}^{ego}$ un perfil de preferencias que induce preferencias estándar que cumplen con las siguientes propiedades:

m_1	m_2	m_3	w_1	w_2
w_1	w_2	m_3	m_2	m_3
w_2	w_1		m_3	m_1
m_1	m_2		m_1	m_2
			w_1	w_2

Asuma que las preferencias de m_3 cumplen con que $\mu' P_{m_3} \mu$, donde μ y μ' se describen a continuación:

$$\mu = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & w_2 & m_3 \end{bmatrix}, \qquad \mu' = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_2 & w_1 & m_3 \end{bmatrix}.$$

Note que μ es el resultado de aplicar AD^e_M. En este caso vemos que m_3 tiene incentivos a reportar preferencias falsas. Si m_3 reporta cualquier $\tilde{R}_{m_3}$ tal que la preferencia estándar cumple con $w_1 \tilde{R}^s_{m_3} m_3$, entonces el resultado de aplicar AD^e_M corresponde a μ'. Por lo tanto AD^e_M no es *strategy-proof* para los miembros de M, ya que $AD^e_M(\tilde{R}_{m_3}, R_{-m_3}) P_{m_3} AD^e_M(R)$.

Por último, considere un nuevo problema de *marriage market* con externalidades. Asuma que $M = \{m_1, m_2, m_3\}$ y $W = \{w_1, w_2\}$. Sea $R \in \mathcal{R}^{ego}$, en donde las preferencias estándar subyacentes cumplen:

m_1	m_2	m_3	w_1	w_2
w_1	w_2	w_1	m_2	m_3
w_2	w_1	m_3	m_3	m_1
m_1	m_2		m_1	m_2
			w_1	w_2

Considere a μ y μ' como los mismos matchings de antes y asuma que $\mu P_{m_3} \mu'$. En este caso el resultado de aplicar AD^e_M es μ'. Se puede apreciar que en este problema μ es un matching individualmente racional. En este caso tenemos que $\mu P_i \mu'$ para todo $i \in M$, lo cual muestra que el resultado de aplicar AD^e_M no es débilmente Pareto eficiente para M. $\square$

En contraposición con el caso sin externalidades, vemos que incluso en un contexto de preferencias egocéntricas se pierden tres importantes propiedades para el caso de *marriage markets*, las cuales están relacionada a la compatibilidad de incentivos como a la eficiencia del mecanismo.

El Teorema 1 muestra que AD^e_M no es un mecanismo débilmente Pareto eficiente para M. Por el Teorema 2 de Roth (1982) sabemos que el mecanismo AD_M es débilmente Pareto eficiente para M. Sin embargo, en un escenario de preferencias egocéntricas es posible encontrar un matching individualmente racional que mejore a todos los miembros de M. Esto se debe a que habrá individuos que mejoren cambiando su pareja y otros que mejoren manteniendo la suya pero mejorando producto del cambio del resto. Efectivamente, se puede ver en la demostración del Teorema 1 que m_1 y m_2 mejoran cambiando sus parejas respecto del resultado de $AD^e_M(R) = \mu'$, mientras que m_3 mejora producto del cambio del resto. En un escenario sin externalidades vemos que no sería posible que μ mejore a todos los miembros de M ya que en ese caso m_3 estaría sin pareja en μ' y en μ.

El Teorema 1 muestra que AD^e_M no es un mecanismo *strategy-proof*. Sabemos que, dado un perfil de preferencias $R \in \mathcal{R}^{ego}$, si R^s es el perfil de preferencias estándar inducido, tenemos que $AD^e_M(R) = AD_M(R^s)$. Como AD_M es un mecanismo *strategy-proof* para M (Dubins y Freedman, 1981), si alguien mejora reportando preferencias falsas es producto de la externalidad que genera sobre otros individuos y no del cambio de su propia pareja. Por lo tanto, el hecho de AD^e_M no sea *strategy-proof* para M está directamente relacionado con el hecho de que el mecanismo AD_M *no* es *non-bossy*. Efectivamente, en la demostración del Teorema 1, m_3 reporta preferencias falsas y con eso consigue que el resultado de aplicar AD^e_M sea distinto al original. En dicho caso se puede observar que m_3 mejora a través de cambiar las parejas del resto y no la suya. En un escenario sin externalidades tendríamos que m_3 no estaría mejor reportando preferencias falsas, pues en μ y en μ' tiene la misma pareja.

Dado que AD^e_M no cumple con ciertas propiedades deseables, como ser Pareto eficiente para M o *strategy-proof* para M, resulta natural preguntarse si existirá algún mecanismo que, en el marco de preferencias egocéntricas, mantenga algunas de las propiedades relacionadas a la estabilidad y a los incentivos.

La búsqueda de mecanismos con buenas propiedades en términos de incentivos resulta difícil incluso en escenarios sin externalidades. Efectivamente, Roth (1982) muestra que no existen mecanismos estables y *strategy-proof*, mientras que Alcalde y Barberà (1994) demuestran que no existen mecanismos Pareto eficientes, individualmente racionales y *strategy-proof*. Si nos enfocamos en la compatibilidad de incentivos de un lado del mercado, Alcalde y Barberà (1994) demuestran que AD_M es el <u>único</u> mecanismo estable y *strategy-proof* para M en la ausencia de externalidades. Lamentablemente, el resultado del Teorema 1—que nos asegura que AD^e_M no es *strategy-proof* para M—nos da pocas esperanzas de encontrar un mecanismo estable que sea compatible con incentivos en estrategias dominantes para M. Desde otra perspectiva, Kojima (2010) muestra que no existen mecanismos que sean estables y *non-bossy* para el caso de *marriage market* sin externalidades. Como en un marco de preferencias egocéntricas la propiedad de *non-bossy* pareciera ser relevante al momento de asegurar *strategy-proof*, el resultado de Kojima (2010) nos da pocas esperanzas de encontrar mecanismos que sean estables y *strategy-proof* para M en contextos de preferencias egocéntricas.

Teorema 2. *Si $\mathcal{R}^{ego} \subseteq \mathcal{R}$, $|M| \geq 3$ y $|W| \geq 2$, entonces no existe un mecanismo $\varphi : \mathcal{R} \to \mathcal{A}(M,W)$ que sea estable y strategy-proof para M.*

Demostración. Es suficiente probar que no existe ningún mecanismo $\varphi : \mathcal{R}^{ego} \to \mathcal{A}(M,W)$ que sea estable y *strategy-proof* para M. Asuma que $M = \{m_1, m_2, m_3, ..., m_a\}$ y $W = \{w_1, w_2, ..., w_b\}$. Suponga que existe un mecanismo estable $\varphi : \mathcal{R}^{ego} \to \mathcal{A}(M,W)$ y sea $R \in \mathcal{R}^{ego}$ un perfil de preferencias que induce *preferencias estándar* que cumplen con las siguientes propiedades:

$$
\begin{array}{ccccc|ccccc}
m_1 & m_2 & m_3 & \cdots & m_a & w_1 & w_2 & w_3 & \cdots & w_b \\
\hline
w_1 & w_2 & m_3 & \cdots & m_a & m_2 & m_3 & w_3 & \cdots & w_b \\
w_2 & w_1 & & & & m_3 & m_1 & & & \\
m_1 & m_2 & & & & m_1 & m_2 & & & \\
& & & & & w_1 & w_2 & & &
\end{array}
$$

Como en todo matching individualmente racional los individuos $m_3, \ldots, m_a, w_3, \ldots, w_b$ quedan solos, no es difícil verificar que en este problema hay solo dos matchings estables:

$$
\mu = \begin{bmatrix} m_1 & m_2 & m_3 & \cdots & m_a & w_3 & \cdots & w_b \\ w_1 & w_2 & m_3 & \cdots & m_a & w_3 & \cdots & w_b \end{bmatrix}, \quad \mu' = \begin{bmatrix} m_1 & m_2 & m_3 & \cdots & m_a & w_3 & \cdots & w_b \\ w_2 & w_1 & m_3 & \cdots & m_a & w_3 & \cdots & w_b \end{bmatrix}.
$$

Asuma que el perfil de preferencias R descrito arriba cumple $\mu' P_{m_3} \mu$. Como φ es un mecanismo estable, el resultado de $\varphi(R)$ será igual a μ o a μ'. Veremos que en ambos casos existirá alguien en M que manipulará sus preferencias para mejorar si situación.

Si $\varphi(R) = \mu$, entonces m_3 tiene incentivos a reportar preferencias falsas. Efectivamente, si m_3 reporta cualquier $\tilde{R}_{m_3}$ tal que la preferencia estándar cumple con $w_1 \tilde{R}^s_{m_3} m_3 \tilde{R}^s_{m_3} i$ para todo $i \in W \backslash \{w_1\}$, entonces μ' es el único matching estable cuando el perfil de preferencias viene dado por $(R_{-m_3}, \tilde{R}_{m_3})$. Por lo tanto, el mecanismo φ no es *strategy-proof* pues $\varphi(R_{-m_3}, \tilde{R}_{m_3}) P_{m_3} \varphi(R)$.

Si $\varphi(R) = \mu'$, entonces m_2 tiene incentivos a reportar preferencias falsas. Efectivamente, si m_2 reporta cualquier $\tilde{R}_{m_2}$ tal que la preferencia estándar cumple con $w_2 \tilde{R}^s_{m_2} m_2 \tilde{R}^s_{m_2} i$ para todo $i \in W \backslash \{w_2\}$, entonces μ es el único matching estable cuando el perfil de preferencias viene dado por $(R_{-m_2}, \tilde{R}_{m_2})$. Por lo tanto, el mecanismo φ no es *strategy-proof* pues $\varphi(R_{-m_2}, \tilde{R}_{m_2}) P_{m_2} \varphi(R)$. $\square$

El Teorema 1 nos aseguraba que AD^e_M no cumplía con ser un estable y *strategy-proof* para M. El Teorema 2 profundiza este resultado de imposibilidad: en todo dominio de perfiles de preferencias que contengan a las preferencias egocéntricas, resultará imposible encontrar un mecanismo que sea estable y *strategy-proof* para M.[12] Recuerde que, en el contexto sin externalidades, no solo sabíamos que AD_M era un mecanismo estable (Gale y Shapley, 1962) y *strategy-proof* para M (Dubins y Freedman, 1981), sino que además, era el único mecanismo que cumplía con dichas propiedades (Alcalde y Barberà, 1994, Teorema 3).

En la demostración del Teorema 2 tenemos un mecanismo estable en el cual la manipulación de preferencias es generada por dos individuos. Uno mejora a través de cambiar su propia pareja y otro mejora cambiando las parejas del resto. Intuitivamente, estamos haciendo uso de la incompatibilidad entre estabilidad y *non-bossy* demostrada por Kojima (2010) en escenarios sin externalidades.

Es importante hacer notar que en la demostración del Teorema 2 solo es necesario que una persona tenga preferencias egocéntricas, la cual en este caso corresponde a m_3. Debido a esto, la imposibilidad

[12] Cuando las preferencias son egocéntricas, tenemos que un matching es estable si y solamente si está en el $\Theta-$core. Una implicancia directa de esto es el hecho de que, dado $\mathcal{R}^{ego} \subseteq \mathcal{R}$, no existe mecanismo $\varphi : \mathcal{R} \to \mathcal{A}(M, W)$ que sea $\Theta-$*core selecting* y *strategy-proof* para M.

de encontrar mecanismos estables y *strategy-proofs* para M solo requiere de una pequeña perturbación respecto del caso sin externalidades.

Por el Teorema 2 sabemos que no existen mecanismos estables y *strategy-proof* para M cuando el dominio de preferencias contiene a las preferencias egocéntricas. En principio, lo anterior no necesariamente implica que no existan mecanismos estables y *strategy-proof* para un conjunto $T \subseteq M$. Sin embargo, al ver la demostración de Teorema 2 podemos notar que no existen mecanismos estables y *strategy-proof* para cualquier $T \subseteq M$ tal que $|T| \geq 2$.

Vale la pena destacar que los resultados anteriores no dependen de la estructura de creencias de los agentes sobre las posibles reacciones del resto cuando ellos desvían para seguir un acuerdo.

A pesar de que el Teorema 2 muestra que no existen mecanismos estables y *strategy proof* para M cuando las preferencias son egocéntricas, puede que existan mecanismos que cumplan con propiedades menos demandantes. En el contexto sin externalidades, por la Proposición 1 de Alcalde y Barberà (1994), sabemos que no existen mecanismos Pareto eficientes, individualmente racionales y *strategy-proof*. Sin embargo en ausencia de externalidades, la literatura conoce un mecanismo que cumple con ser Pareto eficiente para M, individualmente racional para M y *strategy-proof* para M. Este mecanismo corresponde a TTC_M. El siguiente resultado muestra que las propiedades de dicho mecanismo siguen siendo válidas en el marco de preferencias egocéntricas.

Proposición 3.3. *El mecanismo $TTC_M^e : \mathcal{R}^{ego} \to \mathcal{A}(M, W)$ es Pareto eficiente para M, individualmente racional para M y strategy-proof para M.*

Demostración. Ver Apéndice A. □

Debido a la incompatibilidad entre mecanismos estables y *non-bossy* en contextos de *marriage market* sin externalidades (Kojima, 2010), resulta complejo obtener mecanismos con propiedades deseables en presencia de externalidades. A causa de lo anterior, puede ser necesario reducir el dominio de preferencias a través de agregar cierta estructura adicional a las preferencias egocéntricas. Inspirados en Ergin (2002) se presenta la siguiente definición.

Definición 3.2. Dado un problema de *marriage market* con externalidades, en donde $R \in \mathcal{R}^{ego}$, diremos que el perfil de preferencias de W tiene un *ciclo* si existen $w_1, w_2 \in W$ y $m_1, m_2, m_3 \in M$, donde las preferencias estándar de los miembros de W cumplen con

$$m_1 P^s_{w_1} m_2 P^s_{w_1} m_3 P^s_{w_2} m_1.$$

Notar que, a diferencia de Ergin (2002), no se explicitó la *Scarcity Condition*,[13] pues se cumple trivialmente debido a que estamos en un modelo uno-a-uno.

[13] En un contexto de muchos a uno, esta condición hace referencia a que hayan suficientes individuos con una prioridad mayor que m_1, m_2 y m_3, para que estos últimos puedan competir por “cupos” en w_1 y en w_2.

Diremos que un perfil de preferencias egocéntricas de W es *w-acíclico* si no tiene ciclos. Denotaremos por $\mathcal{R}_W^{ego-w}$ al conjunto de todos los perfiles de preferencias egocéntricas de W que son w-acíclicos.

Teorema 3. *Dado un marriage market con externalidades, si las preferencias de los individuos en M están en $\mathcal{R}_M^{ego}$, el conjunto $\mathcal{R}_W^{ego-w}$ es el dominio maximal en $\mathcal{R}_W^{ego}$ donde podemos asegurar lo siguiente*

- *AD_M^e es strategy-proof para M.*
- *AD_M^e es group strategy-proof para M.*
- *AD_M^e es Pareto eficiente para M.*

Demostración. Ver Apéndice A. □

En la demostración de que AD_M^e es *strategy-proof* para M se siguen argumentos análogos a los hechos por Ergin (2002) para probar que, en el marco sin externalidades, la ausencia de ciclos en el perfil de preferencias de W implica que el mecanismo AD_M es *non-bossy* para M. Antes vimos que la única manipulación rentable para los individuos de M, cuando se implementa el mecanismo AD_M^e, es reportar preferencias falsas para cambiar las parejas del resto. Por lo tanto, si el mecanismo AD_M fuera *non-bossy*—y por lo tanto AD_M^e también fuera *non-bossy*—ya no sería posible realizar dicho tipo de manipulaciones.

En la demostración de que AD_M^e es *group strategy-proof* para M se utiliza que AD_M es *group strategy-proof* para M cuando no hay ciclos en el perfil de preferencias de W (Ergin, 2002). Efectivamente, esto nos garantiza que bajo AD_M^e no habrán grupos que mejoren a través de cambiar sus parejas. Sin embargo, en presencia de preferencias egocéntricas, una coalición podría beneficiarse a través de cambiar las parejas del resto manteniendo las suyas. Por suerte, esto último no es posible ya que, en la ausencia de ciclos en el perfil de preferencias de W, el mecanismo AD_M es *group non-bossy* (Afacan, 2011).

La demostración de que AD_M^e es Pareto eficiente para M está relacionada con el Teorema 1 de Ergin (2002), el cual muestra que, en un escenario sin externalidades, la ausencia de ciclos en el perfil de preferencias de W garantiza que AD_M sea Pareto eficiente para M. Como se muestra en la demostración en el Apéndice A, si un matching es Pareto eficiente para M sin externalidades, entonces ese matching es Pareto eficiente para M en el marco de preferencias egocéntricas.

En la demostración del Teorema 2, se puede observar que la existencia de un ciclo en las preferencias de W es primordial para que exista manipulación de preferencias. Por lo tanto, si $|M| \leq 2$, tendremos que AD_M^e siempre será *strategy-proof* para M, *group strategy-proof* para M y Pareto eficiente para M. Esto debido a que no se puede formar un ciclo con dos miembros en M.

Note que si las preferencias de los individuos en M están en $\mathcal{R}_M^{ego}$, el conjunto $\mathcal{R}_W^{ego-w}$ es el dominio maximal en $\mathcal{R}_W^{ego}$ para que el mecanismo AD_M^e cuente con las propiedades descritas en el Teorema 3. Evidentemente, es muy demandante que el perfil de preferencias de W no tenga ciclos, por lo cual el Teorema 3 puede ser visto como un resultado de imposibilidad.

El siguiente ejemplo muestra que, incluso para el dominio de preferencias $\mathcal{R}_M^{ego} \times \mathcal{R}_W^{ego-w}$, el mecanismo AD_M^e no siempre induce un matching estable que sea el mejor para todos los miembros de M.

Ejemplo 3.1. Dado $M = \{m_1, m_2, m_3\}$ y $W = \{w_1, w_2\}$. Sea $R \in \mathcal{R}^{ego-w}$ un perfil de preferencias que induce preferencias estándar que cumplen con las siguientes propiedades:

m_1	m_2	m_3	w_1	w_2
w_2	w_1	m_3	m_1	m_2
w_1	w_2		m_2	m_1
m_1	m_2		m_3	m_3
			w_1	w_2

Note que solo hay dos matchings estables los cuales son:

$$\mu = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_2 & w_1 & m_3 \end{bmatrix}, \qquad \mu' = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & w_2 & m_3 \end{bmatrix}.$$

Asuma que $\mu' P_{m_3} \mu$. En este caso μ es el resultado de aplicar AD_M^e, el cual no es el mejor matching estables para todos los miembros de M, pues m_3 prefiere μ', el cual también es un matching estable.

El ejemplo anterior muestra como incluso realizando hipótesis sobre las preferencias egocéntricas, la estructura de matchings estables difiere del caso sin externalidades.

La Proposición 3.3 muestra que TTC_M es un mecanismo Pareto eficiente para M, individualmente racional para M y *strategy-proof* para M. Motivados por el escenario sin externalidades, podemos restringir el conjunto de las preferencias egocéntricas, con tal de obtener propiedades relacionadas a la estabilidad del mecanismo TTC_M^e. Inspirados en Kesten (2006) damos la siguiente definición.

Definición 3.3. Dado un problema de *marriage market* con externalidades, en donde $R \in \mathcal{R}^{ego}$, diremos que el perfil de preferencias de W tiene un *ciclo débil* si existen $w_1, w_2 \in W$ y $m_1, m_2, m_3 \in M$ tales que, las preferencias estándar de los miembros de W cumplen con $m_1 P_{w_1}^s m_2 P_{w_1}^s m_3$ y al menos una de las siguientes propiedades es satisfecha:

1. $m_3 P_{w_2}^s m_1$.

2. $m_3 P_{w_2}^s m_2$.

Notar que, a diferencia de Kesten (2006), no se utiliza la *Scarcity Condition* pues se cumple trivialmente. Diremos que un perfil de preferencias egocéntricas de W es *w-acíclico fuerte* si no tiene ciclos débiles. Denotaremos por $\mathcal{R}_W^{ego-w-f}$ al conjunto de todos los perfiles de preferencias egocéntricas de W que son w-acíclicos fuertes.

Si las preferencias de los miembros de W tienen un *ciclo fuerte* y cumplen con la condición 1 de la Definición 3.3, entonces tienen un *ciclo* (ver Definición 3.2). Por lo tanto, $\mathcal{R}^{ego-w-f} \subseteq \mathcal{R}^{ego-w}$.

Proposición 3.4. *Dado un marriage market con externalidades, si las preferencias de los individuos en M están en $\mathcal{R}_M^{ego}$, el conjunto $\mathcal{R}_W^{ego-w-f}$ es el dominio maximal en $\mathcal{R}_W^{ego}$ donde podemos asegurar que TTC_M^e es un mecanismo estable.*

Demostración. Ver Apéndice A. □

Sin hacer hipótesis sobre las preferencias egocéntricas, sabíamos que TTC_M^e no era un mecanismo estable. La Proposición 3.4 muestra cual es el dominio maximal en $\mathcal{R}_W^{ego}$ donde sin realizar hipótesis sobre $\mathcal{R}_M^{ego}$ podemos asegurar que TTC_M^e cumpla con dicha propiedad.

4. Estabilidad Cuando las Coaliciones son Extremamente Prudentes

En la Sección 3 se realizaron supuestos sobre la estructura de las externalidades, las cuales tomaron una forma muy particular (preferencias egocéntricas). Otra forma de asegurar que siempre existen soluciones en modelos de *marriage market* con externalidades es realizar supuestos acerca de las creencias de los agentes sobre las posibles reacciones del resto cuando ellos desvían para seguir algún acuerdo. En esta línea, Sasaki y Toda (1996) asumen que las coaliciones son *extremamente prudentes* y con eso obtienen resultados positivos en lo relacionado a la estabilidad.[14] En esta sección desarrollaremos un enfoque similar para ver si los resultados de la Sección 3 siguen siendo válidos en este contexto.

4.1. Definiciones

Al igual que Sasaki y Toda (1996) y Mumcu y Saglam (2010), asumiremos que para todo agente $i \in M \cup W$, la preferencia R_i es completa, transitiva y estricta sobre $\mathcal{A}(M, W)$.

Además, en esta sección asumiremos que toda coalición T es *extremamente prudente* en el sentido que $\Theta^{T,f}(\mu) = \mathcal{A}^{T,f}$ para todo acuerdo $f \in \mathcal{A}(T)$.

Definición 4.1. Dado $m \in M$ e $i \in W \cup \{m\}$, denotemos por $\mu_{i,m}^*$ al matching que cumple con $\mu' P_m \mu_{i,m}^*$ para todo $\mu' \in \mathcal{A}^{\{m,i\},f} \setminus \{\mu_{i,m}^*\}$, con $f(m) = i$. Para cualquier $j, h \in W \cup \{m\}$, definamos la *preferencia aguas abajo de m* como R_m^*, la cual cumple con:

$$j P_m^* h \qquad \Longleftrightarrow \qquad \mu_{j,m}^* \, P_m \, \mu_{h,m}^*.$$

[14] Por coaliciones extremamente prudentes entenderemos que estas consideran todas las posibles desviaciones de otros agentes antes de decidir bloquear un emparejamiento.

Note que, gracias a que R_m es completa, transitiva y estricta, R_m^* es completa, transitiva y estricta. Dado una coalición T, denotaremos por $\mathcal{R}_m^*$ al conjunto de preferencias aguas abajo de m y por $\mathcal{R}_T^*$ al conjunto de perfiles de preferencias aguas abajo de T.

4.2. Mecanismo de Sasaki y Toda (1996)

Al igual que en la sección previa, podemos adaptar AD_M para el caso de preferencias con externalidades, pero tomando en cuenta solo la preferencia aguas abajo de los agentes. A este mecanismo, que fue descrito por Sasaki y Toda (1996), lo llamaremos *algoritmo de aceptación diferida con creencias extremamente prudentes* (AD^*).[15] Lo anterior implica que, para un perfil de preferencias R y el perfil de preferencias aguas abajo inducido R^*, se tendrá

$$AD_M^*(R) = AD_M(R^*).$$

Esto es, las etapas del mecanismo $AD_M^* : \mathcal{R} \to \mathcal{A}(M, W)$ son las siguientes.

- *Etapa 1-a:* Cada $m \in M$ le hace una propuesta al más preferido de los $w \in W \cup \{m\}$ según la preferencia R_m^* si hay alguno aceptable, donde R_m^* es la preferencia aguas abajo obtenida de $R_m \in \mathcal{R}_m$.

- *Etapa 1-b:* Cada $w \in W$ acepta temporalmente a la más atractiva de las propuestas que recibe según la preferencia R_w^*, rechazando las otras. R_w^* es la preferencia aguas abajo obtenida de $R_w \in \mathcal{R}_m$.

- *Etapa k-a:* Cada $m \in M$ que fue rechazado en la etapa previa, le hace una propuesta a la alternativa aceptable más preferida entre aquellas a las cuales aún no le ha hecho propuestas según las preferencia R_m^*. Si no quedan alternativas aceptables, no hace propuestas.

- *Etapa k-b:* Cada $w \in W$ mantiene la alternativa más atractiva entre la que escogió en la etapa previa y las propuestas que recibe en esta etapa, según la preferencia R_w^*.

El algoritmo termina cuando no se hacen más propuestas.

4.3. Resultados

El siguiente resultado fue probado por Sasaki y Toda (1996) y Mumcu y Saglam (2010).

Proposición 4.1. *En un marriage market donde las creencias son extremamente prudentes y las preferencias son estrictas, el mecanismo AD_M^* es estable.*

La Proposición 4.1 muestra que sin realizar ninguna suposición acerca de la estructura de las preferencias, es posible caracterizar un mecanismo estable a partir de realizar hipótesis adicionales

[15] Nos referimos a AD_M^* (AD_W^*) como el mecanismo de aceptación diferida con externalidades cuando M (W) hace las propuestas.

sobre las creencias de los agentes.

Por otro lado, la Proposición 2 de Fonseca-Mairena y Triossi (2019) muestra que, en un contexto de *marriage market* y coaliciones extremamente prudente, si $|M| \geq 2$ y $|W| \geq 2$, entonces no existen mecanismos que sean estables y *strategy-proof* para W.[16] Por la Proposición 4.1 sabemos que AD^*_M es estable, por lo que se concluye directamente que este mecanismo no puede ser *strategy-proof* para M.

En la Sección 3 no se realizaron supuestos acerca de las creencias de los agentes, por lo que algunos resultados de aquella sección seguirán siendo válidos si se impone un tipo particular de creencia. Debido a que $\mathcal{R}^{ego} \subseteq \mathcal{R}$, no es difícil mostrar que los resultados del Teorema 2 siguen siendo aplicables en este contexto. En esta línea, el siguiente resultado muestra de forma paralela a la conclusión del resultado de Fonseca-Mairena y Triossi (2019) (AD^*_M no es *strategy-proof* para M), que la adaptación de AD_M con preferencias aguas abajo también pierde propiedades relevantes respecto del caso sin externalidades.

Teorema 4. *Si $|M| \geq 2$ y $|W| \geq 2$, entonces el mecanismo $AD^*_M : \mathcal{R} \to \mathcal{A}(M,W)$ cumple las siguientes propiedades:*

- *AD^*_M no es strategy-proof para M.*
- *AD^*_M no necesariamente induce el mejor de los matchings estables para los miembros de M, incluso cuando este matching existe.*
- *AD^*_M no es débilmente Pareto eficiente para M entre los matchings estables.*

Demostración. Suponga que $|M| \geq 3$ y $|W| \geq 2$. Por Teorema 2 sabemos que no existen mecanismos, definidos en un dominio que contenga todos los perfiles de preferencias egocéntricas, que sean estables y *strategy-proof* para M. Por la Proposición 4.1 sabemos que AD^*_M es un mecanismo estable. Lo anterior implica que AD^*_M no puede ser *strategy-proof* para M.

Asuma que $M = \{m_1, m_2\}$ y $W = \{w_1, w_2\}$. Sea $R \in \mathcal{R}$ un perfil de preferencias. Los matchings de este problema son los siguientes:

[16] De la demostración de Fonseca-Mairena y Triossi (2019) se desprende directamente que tampoco existen mecanismos que sean estables y *strategy-proof* para M.

$$\mu_1 = \begin{bmatrix} m_1 & m_2 \\ w_1 & w_2 \end{bmatrix}, \qquad \mu_2 = \begin{bmatrix} m_1 & m_2 & w_2 \\ w_1 & m_2 & w_2 \end{bmatrix},$$
$$\mu_3 = \begin{bmatrix} m_1 & m_2 \\ w_2 & w_1 \end{bmatrix}, \qquad \mu_4 = \begin{bmatrix} m_1 & m_2 & w_2 \\ m_1 & w_1 & w_2 \end{bmatrix},$$
$$\mu_5 = \begin{bmatrix} m_1 & m_2 & w_1 \\ m_1 & w_2 & w_1 \end{bmatrix}, \qquad \mu_6 = \begin{bmatrix} m_1 & m_2 & w_1 \\ w_2 & m_2 & w_1 \end{bmatrix},$$
$$\mu_7 = \begin{bmatrix} m_1 & m_2 & w_1 & w_2 \\ m_1 & m_2 & w_1 & w_2 \end{bmatrix}.$$

Las preferencias de los agentes son:

m_1	m_2	w_1	w_2
μ_1	μ_1	μ_1	μ_1
μ_5	μ_5	μ_5	μ_5
μ_3	μ_3	μ_3	μ_4
μ_4	μ_2	μ_6	μ_3
μ_7	μ_6	μ_2	μ_2
μ_2	μ_4	μ_4	μ_3
μ_6	μ_7	μ_7	μ_7

Podemos notar que las preferencias aguas abajo de los agentes corresponden a:

m_1	m_2	w_1	w_2
m_1	w_2	m_1	m_2
w_1	w_1	m_2	m_1
w_2	m_2	w_1	w_2

En este caso, tenemos que $AD^*_M(R) = \mu_5$. En primer lugar, se puede notar que μ_1 es un matching estable, el cual mejora a todos los miembros de M. Por lo tanto, AD^*_M no es débilmente Pareto eficiente para M entre los matchings estables y no siempre induce el mejor de los matchings estables para M, incluso cuando dicho matching existe.

Por otro lado, si m_1 reporta las preferencias

$$\tilde{R}_{m_1} : \mu_1 P_{m_1} \mu_5 P_{m_1} \mu_3 P_{m_1} \mu_4 P_{m_1} \mu_2 P_{m_1} \mu_6 P_{m_1} \mu_7$$

tenemos que $AD^*_M(\tilde{R}_{m_1}, R_{-m_1}) = \mu_1$, lo cual mejorara todos y en particular a m_1. En consecuencia AD^*_M no es un mecanismo *strategy-proof* para M. □

Sigue de la Sección 3 que, si las preferencias de los miembros de W son egocéntricas y w-acíclicas (ver Definición 3.2), entonces el mecanismo AD^e_M es *strategy-proof*, *group strategy-proof* y Pareto eficiente para los miembros del grupo M. Una importante implicación de la demostración del Teorema 4, es que no es claro como restringir las preferencias de los miembros de W para asegurar que el mecanismo AD^*_M cumpla estas tres propiedades. Efectivamente, cualquier definición de aciclicidad debería ser satisfecha en un mercado en que $|M| = 2$. Sin embargo, el ejemplo dado en la demostración del Teorema 4 muestra que, en un *marriage market* con dos individuos en cada lado del mercado, el algoritmo AD_M^* no cumple ninguna de las tres propiedades antes descritas.

5. Preferencias Top y Preferencias Dominantes

Como se ha mencionado anteriormente, sin hacer supuestos acerca de la estructura de las externalidades o de las creencias de los agentes, el $\Theta-$*core* puede ser vacío (Sasaki y Toda, 1996). El objetivo de esta sección es proponer dos estructuras sobre las preferencia, las cuales puedan ser aplicadas a diversos escenarios.

En ciertos contextos la población tiende a segmentarse en grupos. Ejemplos de aquello son el mercado educacional con estudiantes y colegios, la provisión de salud privada con pacientes y hospitales o el mercado laboral con empresas y trabajadores. Debido a ello, puede que las personas pertenecientes a un grupo les importe sobre todo pertenecer de una forma en particular a dicho grupo, independiente de lo que haga el resto. Esto es, podría existir cierta cohesión entre los miembros de dichos grupos, lo cual podría tener efectos en la existencia de matchings estables o en el $\Theta-$*core*. En el siguiente apartado se expondrá dos tipos de preferencias que formalicen la intuición antes descrita.

5.1. Definiciones

Asumiremos que $M \cup W$ puede ser *particionado* en las coaliciones $S_1, \ldots, S_n$. Diremos que índice del agente i es τ si y solamente si $i \in S_\tau$.

Definición 5.1. El conjunto S_τ tiene un *acuerdo top* si existe $\bar{f}_\tau \in \mathcal{A}(S_\tau)$, al cual llamaremos *acuerdo top de* S_τ, tal que todos los miembros de S_τ prefieren estar en matchings donde se implementa $\bar{f}_\tau$. Esto es, para todo $i \in S_\tau$ se cumple la siguiente propiedad:

$$\mu P_i \mu' \qquad \forall \mu \in \mathcal{A}^{S_\tau, \bar{f}_\tau}, \quad \forall \mu' \in \mathcal{A}(M, W) \setminus \mathcal{A}^{S_\tau, \bar{f}_\tau}.$$

Diremos que el perfil de preferencias $R = (R_i)_{i \in M \cup W}$ es *top* si cada conjunto S_τ tiene un acuerdo top. Denotaremos a $\mathcal{R}^t$ como al conjunto de todos los perfiles de preferencias top.

Dado $i \in S_\tau$, sea $\mathcal{D}_i = \{\mu \in \mathcal{A}(M, W) : \exists \sigma < \tau, \mu(i) \in S_\sigma\}$. Esto es, $\mathcal{D}_i$ es el conjunto de matchings en donde el agente i está emparejado con alguien en $\bigcup_{\sigma < \tau} S_\sigma$. Se subentiende que si $i \in S_1$, entonces $\mathcal{D}_i = \emptyset$.

Definición 5.2. El conjunto S_τ tiene un *acuerdo dominante* si existe $\widehat{f}_\tau \in \mathcal{A}(S_\tau)$, al cual llamaremos *acuerdo dominante de* S_τ, tal que todos sus miembros prefieren estar en matchings donde se implementa $\widehat{f}_\tau$ o en matchings con personas de un índice menor. Esto es, para todo $i \in S_\tau$ se cumple la siguiente propiedad:

$$\mu P_i \mu' \qquad \forall \mu \in \mathcal{A}^{S_\tau, \widehat{f}_\tau} \cup \mathcal{D}_i, \quad \forall \mu' \in \mathcal{A}(M,W) \setminus (\mathcal{A}^{S_\tau, \widehat{f}_\tau} \cup \mathcal{D}_i).$$

Diremos que el perfil de preferencias $R = (R_i)_{i \in M \cup W}$ es *dominante* si cada conjunto S_τ tiene un acuerdo dominante. Denotaremos a $\mathcal{R}^d$ como al conjunto de todos los perfiles de preferencias dominantes.

Teorema 5. *Si $R \in \mathcal{R}^t \cup \mathcal{R}^d$, entonces existe un único matching en el $\Theta-$core.*

Demostración. Ver Apéndice A. □

En el contexto del Teorema 5, el único matching que está en el $\Theta-$*core* es aquel en el cual todos los miembros de todas las coaliciones siguen el acuerdo top o el acuerdo dominante, según corresponda. Podemos notar que el resultado del Teorema 5 es independiente de las creencias de los agentes.

Por Roth y Sotomayor (1990) sabemos que en ausencia de externalidades el conjunto de matchings estables coincide con el conjunto de matchings en el *core*. Por la Proposición 3.1 sabemos que el resultado anterior sigue siendo válido en un contexto de preferencias egocéntricas. Lo anterior no resulta cierto al incorporar otro tipo de externalidades. Sasaki y Toda (1996) muestran un ejemplo en el cual, asumiendo que los individuos son extremamente prudentes, el $\Theta-$*core* es vacío y difieren del conjunto de matchings estables.[17] El siguiente resultado muestra que un marco de preferencias top o preferencias dominantes, el conjunto de matchings estables difiere del conjunto de matchings en el $\Theta-$*core*, incluso si este tiene un único elemento y las coaliciones son extremamente prudentes.

Teorema 6. *Suponga que las coaliciones son extremamente prudentes. Si $R \in \mathcal{R}^t \cup \mathcal{R}^d$, entonces el conjunto de matchings estables puede contener de manera estricta al conjunto de matching en el $\Theta-$core, incluso cuando el $\Theta-$core tiene un único elemento.*

Demostración. Sea $M = \{m_1, m_2, m_3\}$, $W = \{w_1, w_2,\}$ y $R \in \mathcal{R}^t \cup \mathcal{R}^d$. En donde se tendrá que $S_1 = \{m_1, w_1\}$ y $S_2 = \{m_2, m_3, w_2\}$. Considere los siguientes matchings:

$$\mu = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & w_2 & m_3 \end{bmatrix}, \qquad \mu' = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & m_2 & w_2 \end{bmatrix}.$$

Asuma que el acuerdo top (o dominante) de S_1 corresponde a $f_1(m_1) = w_1$ y el acuerdo top (o dominante) de S_2 corresponde a $f_2(m_2) = w_2$ y $f_2(m_3) = m_3$. Por el Teorema 5 sabemos que μ es el único matching en el $\Theta-$*core*, por lo que también es estable.[18] Asuma que, para todo $i \in M \cup W$, se cumple que $\mu P_i \mu' P_i \eta$ para todo $\eta \in \mathcal{A}(M,W) \setminus \{\mu, \mu'\}$. Mostraremos que μ' también es un matching estable. Supongamos que una coalición T de uno o dos individuos bloquea a μ' para seguir un acuerdo f. Se puede notar que $\Theta^{T,f}(\mu') \cap \mathcal{A}(M,W) \setminus \{\mu, \mu'\} \neq \emptyset$, esto debido a que las coaliciones son

[17] Si las coaliciones son extremamente prudentes, el conjunto de matchings estables es diferente de vacío (ver Proposición 4.1). Recodemos que si una coalición T es extremamente prudente si y solamente si $\Theta^{T,f}(\mu) = \mathcal{A}^{S,f}$ para todo $f \in \mathcal{A}(T)$.

[18] Si un matching está en el $\Theta-$*core*, entonces es estable.

extremamente prudentes. En ese caso existe un matching en $\Theta^{T,f}(\mu')$ que es peor que μ' para todos los miembros de T. Lo anterior implica que toda coalición T de uno o dos individuos no bloquearía a μ', por lo que μ' es estable. $\square$

El ejemplo anterior ilustra que a pesar de que el $\Theta-$*core* sea único en presencia de preferencias top o dominante, el conjunto de matchings estables puede ser mayor. Preguntarnos sobre qué hipótesis adicionales se deben realizar si se busca obtener un único matching estable, motiva la siguiente proposición.

Proposición 5.1. *Si $R \in \mathcal{R}^t \cup \mathcal{R}^d$ y $|S_\tau| \leq 2$, para todo $\tau \in \{1, 2, \ldots, n\}$, entonces existe un único matching estable.*

Demostración. Sea $S_1, \ldots, S_n$ las n coaliciones disjuntas de $M \cup W$. Además f_τ, con $\tau \in \{1, 2, ..., n\}$, corresponde al acuerdo que es top o dominante para los miembros S_τ. Por el Teorema 5 sabemos que existe al menos un matching estable, el cual corresponde a $\bigcap_{\tau=1}^{n} \mathcal{A}^{S_\tau, f_\tau} = \mu$. Mostraremos que no hay otro matching estable. Supongamos que existe un $\mu' \neq \mu$, que también es estable. Si $\mu' \neq \mu$, entonces $T = \{i \in M \cup W : \mu'(i) \neq \mu(i)\} \neq \emptyset$. Considere a $i' \in T$ como la persona con el menor índice entre los miembros de T. Sea j' el índice de i'. Observemos que, como en $S_{j'}$ están las personas con menor índice en T, tenemos que $\mu' \in \mathcal{A}(M, W) \setminus \mathcal{A}^{S_j, f_j} \cup \mathcal{D}_i$ para todo $i \in S_{j'}$. Notemos que el par $\{i', \mu(i')\}$ bloquean μ'. Esto, debido a la definición de preferencia top o dominante, en la cual tenemos que $\eta P_i \mu'$ para todo $\eta \in \Theta^{\{i', \mu(i')\}, f_{j'}}(\mu')$. $\square$

La Proposición 5.1 muestra una condición suficiente para que exista un único matching estable cuando el perfil de preferencias es top o dominante. Lo anterior implica que el tamaño de las coaliciones incide en el número de matchings estables. En ese sentido, podemos ver como en la demostración del Teorema 6 la coalición S_2 tiene tres miembros y hay más de un matching estable.

6. Sobre el *Bargaining Set* con Externalidades

Cuando el $\Theta-$*core* es vacío, todo matching es bloqueado por al menos una coalición. Podría ser que algunos miembros de un grupo que bloquean un matching hubieran preferido bloquear con otra coalición. En ese caso el bloqueo original podría no ser creíble, pues podrían existir agentes con incentivos a romper el acuerdo pactado con el primer grupo que bloqueó. Las ideas anteriores dan paso a una definición menos fuerte que estabilidad: el *bargaining set*.

Klijn y Massó (2003) adaptan el concepto de *bargaining set* de Zhou (1994) al contexto de un *marriage market sin* externalidades. Ellos demuestran que, cuando las preferencias son estrictas, el *bargaining set* coincide con el conjunto de matchings que son débilmente estables y débilmente Pareto eficientes.[19] Para el caso con externalidades, hasta nuestro conocimiento, Hafalir (2008) ha sido el único que ha trabajado con el *bargaining set*. En la subsección siguiente, se introducirá de manera formal el *bargaining set* de Hafalir (2008), así como una nueva propuesta para este concepto.

[19] Un par de bloqueo (m, w) es *débil* si existe una $w' \in W$ o un $m' \in M$ que cumpla con $w' P^s_m w$ o $m' P^s_w m$. Diremos que un matching es *débilmente estable* si todos los pares de bloqueo son débiles.

6.1. Modelo

Considere un *marriage market problem*

$$\Lambda = \left\langle M, W, (R_i)_{i \in M \cup W}, \left(\Theta^{T,f}\right)_{T \subseteq M \cup W, f \in \mathcal{A}(T)} \right\rangle$$

en el cual se cumplen las siguientes propiedades: $|M| = |W|$, las coaliciones tienen el mismo número de miembros de M y W, para todo $i \in M \cup W$ las preferencias R_i sobre $\mathcal{A}(M, W)$ son estrictas, las coaliciones son extremamente prudentes y cada conjunto $\mathcal{A}(T)$ solo considera matchings donde nadie queda solo.[20]

Primero adaptaremos la definición usada por Hafalir (2008) a nuestra notación. Por conveniencia, diremos que (T, f) es un *μ-bloqueo* si la coalición T consigue bloquear el matching μ a través del acuerdo f.[21] Además, dado $T' \subseteq T$ diremos que (T, f) es un *μ-bloqueo para T'* si se cumplen los siguientes requerimientos:

- $\eta R_i \mu$ para todo $\eta \in \Theta^{T,f}(\mu)$ y para todo $i \in T'$.
- Para todo $\eta \in \Theta^{T,f}(\mu)$ existe un $i \in T'$ que $\eta P_i \mu$.

Definición 6.1. *[Hafalir (2008, Definición 3)]*
Una coalición S *contra-bloquea en el sentido de Hafalir* un μ-bloqueo (T, f) si $T \cap S \neq \emptyset$ y existe $g \in \mathcal{A}(S)$ tal que:

1. (S, g) es un μ-bloqueo.
2. Para todo $i \in S \cap T$ existe $\eta \in \Theta^{T,f}(\mu)$ tal que (S, g) es un η-bloqueo para i.

Un matching μ está en el *bargaining set a la Hafalir*, denotado por $\mathcal{B}^H$, si todo μ-bloqueo es contra-bloqueado en el sentido de Hafalir.

Ahora presentaremos la definición de bargaining set propuesta por nosotros.

Definición 6.2.
Una coalición S *contra-bloquea* un μ-bloqueo (T, f) si $T \cap S \neq \emptyset$ y existe $g \in \mathcal{A}(S)$ tal que:

1. (S, g) es un μ-bloqueo.
2. Para todo $i \in S \cap T$ y $\eta \in \Theta^{T,f}(\mu)$, (S, g) es un η-bloqueo para i.

Un matching μ está en el *bargaining set*, denotado por $\mathcal{B}$, si todo μ-bloqueo es contra-bloqueado.

[20] La mayoría de estas propiedades no tiene mayor efecto en las definiciones siguientes, sin embargo permiten estudiar un escenario homologo al desarrollado por Hafalir (2008).

[21] El significado de bloqueo viene de la Definición 2.1.

Note que, para *contra-bloquear en el sentido de Hafalir* es suficiente que, al seguir el bloqueo alternativo (S, g), cada individuo $i \in S \cap T$ mejore en al menos un escenario resultante del bloqueo original. Por otra parte, para *contra-bloquear*, al seguir el bloqueo alternativo (S, g) los individuos pertenecientes a ambas coaliciones deben mejorar en todos los posibles escenarios resultantes del bloque original. Lo anterior implica que *contra-bloquear* es más difícil que *contra-bloquear en el sentido de Hafalir*. Por lo tanto, nuestro concepto de *bargaining set* es un refinamiento del concepto propuesto por Hafalir (2008).

Hafalir (2008, Proposición 4) demostró que, si las preferencias son tales que existe un matching que es el peor para todos los individuos, entonces $\mathcal{B}^H$ es diferente de vacío. A continuación mostramos que, si para las coaliciones es más difícil contra-bloquear (en el sentido de la Definición 6.2), el resultado de existencia de Hafalir (2008) deja de ser válido.

Proposición 6.1. *El bargaining set $\mathcal{B}$ está contenido en $\mathcal{B}^H$. Además, si los individuos son extremamente prudentes, $\mathcal{B}$ puede ser vacío incluso si hay un matching que es el peor para todos.*

Demostración. El hecho de que $\mathcal{B} \subseteq \mathcal{B}^H$ viene de la Definición 6.1 y 6.2. Por lo tanto, mostraremos que, cuando los individuos son extremamente prudentes, el conjunto $\mathcal{B}$ puede ser vacío incluso si hay un matching que es el peor para todos. Basado en Hafalir (2008), asuma que $M = \{m_1, m_2, m_3\}$ y $W = \{w_1, w_2, w_3\}$. Los matchings son:

$$\mu_1 = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \end{bmatrix}, \quad \mu_3 = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \end{bmatrix},$$

$$\mu_4 = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \end{bmatrix}, \quad \mu_5 = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_3 & w_2 & w_1 \end{bmatrix}, \quad \mu_6 = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \end{bmatrix}.$$

Las preferencias son

m_1	m_2	m_3	w_1	w_2	w_3
μ_5	μ_4	μ_3	μ_1	μ_3	μ_3
μ_2	μ_3	μ_2	μ_2	μ_5	μ_5
μ_4	μ_5	μ_4	μ_4	μ_2	μ_2
μ_1	μ_2	μ_1	μ_3	μ_4	μ_4
μ_3	μ_1	μ_5	μ_5	μ_1	μ_1
μ_6	μ_6	μ_6	μ_6	μ_6	μ_6

En un ejemplo similar, Hafalir (2008) muestra que $\{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5\} \notin \mathcal{B}^H$. Como $\mathcal{B} \subseteq \mathcal{B}^H$ podemos concluir que $\{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5\} \notin \mathcal{B}$.

Note que, μ_6 es bloqueado por el par (m_1, w_3) pues μ_3 y μ_5 son los únicos matchings que se pueden formar cuando m_1 y w_3 bloquean y se tiene que $\mu_3 P_i \mu_6$ y $\mu_5 P_i \mu_6$ con $i \in \{m_1, w_3\}$. Como fue demostrado en el Ejemplo 4 de Hafalir (2008), la única coalición de bloqueo para μ_5 es $\{m_2, w_1, m_3, w_2\}$. Como m_1 y w_3 no están presentes en esa coalición, no hay forma de contra-bloquear. Por lo tanto, como la definición de *bargaining set* requiere que todo bloqueo sea contra-bloqueado, llegamos a que $\mathcal{B}$ es vacío. $\square$

7. Conclusión

En presencia de externalidades puede que no existan matchings estables para problemas de *marriage markets*. Por esta razón, para asegurar la existencia de soluciones estables, en la literatura reciente se han propuesto diversos modelos con externalidades que imponen restricciones sobre las preferencias de los individuos, o sobre las creencias que ellos tienen al momento de evaluar un desvío a partir de un matching (ver, por ejemplo, Sasaki y Toda (1996), Dutta y Massó (1997) y Mumcu y Saglam (2010)). En este trabajo estudiamos los efectos que esas restricciones tienen sobre los incentivos de los individuos al momento de revelar sus preferencias. Además, proponemos restricciones alternativas sobre las preferencias para asegurar la existencia de soluciones estables.

En un primer modelo, asumiendo que las preferencias son egocéntricas—como en el modelo habitacional de Hong y Park (2022)—mostramos que un algoritmo análogo a aceptación diferida (AD_M^e) nos permite encontrar matchings estables. A pesar de lo poco relevantes que son las externalidades asociadas a preferencias egocéntricas, pues los individuos se interesan en primera instancia por su propia pareja, el mecanismo AD_M^e pierde importantes propiedades que aceptación diferida cumple en el modelo clásico. Efectivamente, AD_M^e no es *strategy-proof* para M ni débilmente Pareto eficiente para M (ver Teorema 1). Intuitivamente, estos resultados negativos se deben a que el algoritmo propuesto—al igual que aceptación diferida—*no* es *non-bossy*. De hecho, en este contexto los efectos de las externalidades en los incentivos son mucho más profundos: no existen mecanismos que sean estables y *strategy-proof* para M cuando el dominio de perfiles de preferencias contiene a las preferencias egocéntricas (ver Teorema 2). A pesar de este resultado de imposibilidad, desde la perspectiva de los individuos del conjunto M, demostramos que el algoritmo *top trading cycle* continua siendo Pareto eficiente, individualmente racional y *strategy-proof* (ver Proposición 3.3). Por otro lado, inspirados por los resultados de Ergin (2002) y Kesten (2006) para modelos sin externalidades, determinamos el dominio maximal sobre los perfiles de preferencias de W en el cual, sin suponer restricciones adicionales sobre las preferencias de los miembros de M, el mecanismo AD_M^e cumple con ser *group strategy-proof* y Pareto eficiente para los miembros del grupo M (ver Teorema 3).

En un segundo modelo, sin realizar supuestos acerca de la estructura de las preferencias, analizamos los incentivos que individuos extremamente prudentes tienen para reportar sus verdaderas preferencias. En este contexto, Sasaki y Toda (1996) probaron que un mecanismo similar a aceptación diferida (AD_M^*) es estable. Nosotros demostramos que los resultados de imposibilidad que fueron probados para el modelo con preferencias egocéntricas siguen siendo válidos en este contexto. Esto es, para los individuos del conjunto M, el mecanismo AD_M^* no es *strategy-proof* ni débilmente Pareto eficiente (ver Teorema 4). El hecho que AD_M^* no es *strategy-proof* para M fue previamente probado por Fonseca-Mairena y Triossi (2019).

Como alternativa a los modelos antes descritos, se introdujeron dos nuevas estructuras de preferencias llamadas *preferencias top* y *preferencias dominantes*. Por un lado, demostramos que si un perfil de preferencias pertenece a cualquiera de estos dominios, entonces el $\Theta-$*core* tiene un único elemento (ver Teorema 5). También se demostró que para las preferencias usadas en esta sección, el conjunto de matchings en el $\Theta-core$ puede estar estrictamente contenido en el conjunto de matchings estables (ver Teorema 6). Esto muestra que el resultado clásico de Roth y Sotomayor (1990)—la equivalencia entre el núcleo y estabilidad—se pierde en presencia de externalidades.

Finalmente, propusimos una definición para el *bargaining set* en la presencia de externalidades, la cual es un refinamiento del concepto análogo propuesto por Hafalir (2008). Esencialmente, asumimos que las coaliciones son más prudentes al momento de contra-bloquear. En este contexto, mostramos que nuestro *bargaining set* puede ser vacío incluso cuando se cumplen las condiciones suficientes para que el *bargaining ser* de Hafalir (2008) tenga elementos (ver Proposición 6.1).

Referencias

Abdulkadiroğlu, A., y Sönmez, T. (2003, June). School choice: A mechanism design approach. *American Economic Review*, *93*(3), 729-747.

Afacan, M. O. (2011). On the'group non-bossiness' property. *Available at SSRN 1854103*.

Alcalde, J., y Barberà, S. (1994). Top dominance and the possibility of strategy-proof stable solutions to matching problems. *Economic Theory*, *4*(3), 417–435.

Aziz, H., y Lee, E. (2018). The temporary exchange problem. *arXiv preprint arXiv:1807.05514*.

Braitt, M. d. S., y Torres-Martínez, J. P. (2021). Matching with externalities: The role of prudence and social connectedness in stability. *Journal of Mathematical Economics*, *92*, 95-102.

Contreras, J. L., y Torres-Martínez, J. P. (2021). The roommate problem with externalities. *International Journal of Game Theory*, *50*(1), 149–165.

Dubins, L. E., y Freedman, D. A. (1981). Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly*, *88*(7), 485–494.

Dutta, B., y Massó, J. (1997). Stability of matchings when individuals have preferences over colleagues. *Journal of Economic Theory*, *75*(2), 464-475.

Ergin, H. I. (2002). Efficient resource allocation on the basis of priorities. *Econometrica*, *70*(6), 2489–2497.

Fonseca-Mairena, M. H., y Triossi, M. (2019). Incentives and implementation in marriage markets with externalities. *Economics Letters*, *185*, 108688.

Fonseca-Mairena, M. H., y Triossi, M. (2022). Incentives and implementation in allocation problems with externalities. *Journal of Mathematical Economics*, *99*, 102613.

Gale, D., y Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

Graziano, M. G., Meo, C., y Yannelis, N. C. (2020). Shapley and scarf housing markets with consumption externalities. *Journal of Public Economic Theory*, *22*(5), 1481–1514.

Hafalir, I. E. (2008). Stability of marriage with externalities. *International Journal of Game Theory*, *37*(3), 353–369.

Hong, M., y Park, J. (2022). Core and top trading cycles in a market with indivisible goods and externalities. *Journal of Mathematical Economics*, *100*, 102627.

Kesten, O. (2006). On two competing mechanisms for priority-based allocation problems. *Journal of Economic Theory*, *127*(1), 155-171.

Klaus, B. (2021). Characterizing the top trading cycles rule for housing markets with lexicographic preferences when externalities are limited.

Klaus, B., y Meo, C. (2021). *The core for housing markets with limited externalities* (Inf. Téc.). Université de Lausanne, Faculté des HEC, Département d'économie.

Klijn, F., y Massó, J. (2003). Weak stability and a bargaining set for the marriage model. *Games and Economic Behavior*, *42*(1), 91-100.

Knuth, D. E. (1976). *Mariages stables et leurs relations avec d'autres problemes combinatoires: introduction a l'analysis mathematique des algorithmes-*. Les Presses de l'Université de Montréal.

Kojima, F. (2010). Impossibility of stable and nonbossy matching mechanisms. *Economics Letters*, *107*(1), 69-70.

Li, S. (1993). *Compititive matching equilibrium and multiple principal-agent models* (Inf. Téc.). Minnesota-Center for Economic Research.

Mumcu, A., y Saglam, I. (2007). The core of a housing market with externalities. *Economics Bulletin*, *3*(55), 1–5.

Mumcu, A., y Saglam, I. (2010). Stable one-to-one matchings with externalities. *Mathematical Social Sciences*, *60*(2), 154-159.

Pápai, S. (2000). Strategyproof assignment by hierarchical exchange. *Econometrica*, *68*(6), 1403–1433.

Piazza, A., y Torres-Martínez, J. P. (2021). Coalitional stability in matching problems with externalities and random preferences. *Available at SSRN 3965063*.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of operations research*, *7*(4), 617–628.

Roth, A. E., y Sotomayor, M. A. O. (1990). *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press.

Sasaki, H., y Toda, M. (1996). Two-sided matching problems with externalities. *Journal of Economic Theory*, *70*(1), 93-108.

Shapley, L., y Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, *1*(1), 23-37.

Zhou, L. (1994). A new bargaining set of an n-person game and endogenous coalition formation. *Games and Economic Behavior*, *6*(3), 512-526.

Apéndice A

Proposición 3.1

Demostración. Primero mostraremos que $\mathcal{C}(R) = \mathcal{S}(R)$. Por definición sabemos que $\mathcal{C}(R) \subseteq \mathcal{S}(R)$. Mostraremos que $\mathcal{S}(R) \subseteq \mathcal{C}(R)$. Supongamos que esto no es cierto. En ese caso existe un matching $\mu \in \mathcal{S}(R)$ pero que $\mu \notin \mathcal{C}(R)$. Esto último implica que existe una coalición T y un acuerdo $f \in \mathcal{A}(T)$ en donde se tiene:

1. $\mu(i) \neq f(i)$, para algún $i \in T$.
2. $\eta R_i \mu$ para todo $\eta \in \Theta^{T,f}(\mu)$ y todo $i \in T$.
3. Para todo $\eta \in \Theta^{T,f}(\mu)$ existe un $i \in T$ que $\eta P_i \mu$.

Por 1 aprendemos que $T' = \{i \in T : f(i) \neq \mu(i)\}$ es diferente de vacío. Sea i alguien en T'. Como $f \in \mathcal{A}(T)$, entonces existe un $j \in T'$ en donde $j = f(i)$. Podemos notar que se tiene que $jP_i^s\mu(i)$ y $iP_j^s\mu(j)$. Caso contrario se tendría que $\mu(i)P_ij$ y $\mu(j)P_ji$. Por la definición de preferencias egocéntricas se desprendería que $\mu P_i \eta$ y $\mu P_j \eta$ para todo η en donde $\eta(j) = i$, lo cual contradice 2. Dado que se tiene $jP_i^s\mu(i)$ y $iP_j^s\mu(j)$, entonces tendríamos que j e i bloquearían μ.[22] Lo cual contradice la estabilidad de μ.

Mostremos ahora que $\mathcal{S}(R) = \mathcal{S}(R^s)$. Primero mostremos que $\mathcal{S}(R) \subseteq \mathcal{S}(R^s)$. Supongamos que esto no es cierto. En ese caso existe un matching $\mu \in \mathcal{S}(R)$, pero que $\mu \notin \mathcal{S}(R^s)$. Esto último implica que existe una coalición T de uno o dos individuos y un acuerdo $f \in \mathcal{A}(T)$ en donde se tiene:

1. $\mu(i) \neq f(i)$, para algún $i \in T$.
2. $f(i)R_i^s\mu(i)$ todo $i \in T$.
3. Existe un $i \in T$ que $f(i)P_i^s\mu(i)$.

[22] Se puede notar que la demostración sigue siendo válida si j e i son el mismo individuo.

Por 1 aprendemos que $T' = \{i \in T : f(i) \neq \mu(i)\}$ es diferente de vacío. Como las preferencias estándar son estrictas, podemos notar que $f(i)P_i^s\mu(i)$ para todo $i \in T'$. Notemos que en ese caso se cumple que $\eta P_i \mu$ para todo $\eta \in \Theta^{T',f'}(\mu)$ y todo $i \in T'$, en donde f' corresponde al acuerdo f para los miembros de T'. Lo anterior implica que la coalición T' bloquearía a μ cuando las preferencias son egocéntricas, lo cual es una contradicción con la estabilidad de μ cuando el perfil de preferencias es un perfil de preferencias egocéntricas.

Ahora mostraremos que $\mathcal{S}(R^s) \subseteq \mathcal{S}(R)$. Supongamos que esto no es cierto. En ese caso existe un matching $\mu \in \mathcal{S}(R^s)$, pero que $\mu \notin \mathcal{S}(R)$. Esto último implica que existe una coalición T de uno o dos individuos y un acuerdo $f \in \mathcal{A}(T)$ en donde se tiene:

1. $\mu(i) \neq f(i)$, para algún $i \in T$.
2. $\eta R_i \mu$ para todo $\eta \in \Theta^{T,f}(\mu)$ y todo $i \in T$.
3. Para todo $\eta \in \Theta^{T,f}(\mu)$ existe un $i \in T$ que $\eta P_i \mu$.

Por 1 aprendemos que $T' = \{i \in T : f(i) \neq \mu(i)\}$ es diferente de vacío. Sea i alguien en T'. Como $f \in \mathcal{A}(T)$, entonces existe un $j \in T'$ en donde $j = f(i)$. Podemos notar que se tiene que $jP_i^s\mu(i)$ y $iP_j^s\mu(j)$. Caso contrario se tendría que $\mu(i)P_ij$ y $\mu(j)P_ji$. Por la definición de preferencias egocéntricas se desprendería que $\mu P_i \eta$ y $\mu P_j \eta$ para todo η en donde $\eta(j) = i$, lo cual contradice 2. Dicho aquello, se se tiene que $jP_i^s\mu(i)$ y que $iP_j^s\mu(j)$, entonces tendríamos que j e i bloquearían μ. Lo cual contradice la estabilidad de dicho matching cuando las preferencias son las preferencias estándar inducidas de R.

Por Roth y Sotomayor (1990) sabemos que para todo $R^s \in \mathcal{R}^s$ se cumple que $\mathcal{S}(R^s) = \mathcal{C}(R^s)$. Juntando los resultados anteriores concluimos que para todo $R \in \mathcal{R}^{ego}$ y R^s es el perfil de preferencias estándar inducido de R, tenemos que $\mathcal{C}(R) = \mathcal{S}(R) = \mathcal{S}(R^s) = \mathcal{C}(R^s)$.

Como para todo $R \in \mathcal{R}^{ego}$ y R^s el perfil de preferencias estándar inducido se tiene que $AD_M(R^s) = AD_M^e(R)$. También sabemos por Gale y Shapley (1962) que $AD_M(R^s) \in \mathcal{S}(R^s)$. Juntando los resultados anteriores concluimos que para todo $R \in \mathcal{R}^{ego}$ se tiene que $AD_M(R^s) \in \mathcal{C}(R)$, por lo que el mecanismo AD_M^* es $\Theta-$*core selecting*. $\square$

Proposición 3.3

Demostración. Primero mostraremos que $TTC_M^e : \mathcal{R}^{ego} \to \mathcal{A}(M, W)$ es Pareto eficiente para M. Supongamos que esto no es cierto. En ese caso entonces existe un perfil de preferencias $R \in \mathcal{R}^{ego}$, en donde el matching $TTC_M^e(R) = \mu$ no es Pareto eficiente para M. En ese caso existe un matching μ' que cumple con:

1. $\mu' R_i \mu$ para todo $i \in M$.
2. $\mu' P_i \mu$ para algún $i \in M$.

Sea $T = \{i \in M : \mu'(i) \neq \mu(i)\}$. Aprendemos que $T \neq \emptyset$, caso contrario se contradeciría 2. Sea $R^s = (R_i)_{i \in M \cup W}$ el perfil de preferencias estándar inducido de R y R_i^s la preferencia estándar del individuo i. Como las preferencias son egocéntricas tenemos que las preferencias estándar cumplen con $\mu'(i) P_i^s \mu(i)$ para todo $i \in T$. Caso contrario se tendría que $\mu(i) R_i^s \mu'(i)$, pero como $\mu(i) \neq \mu'(i)$ implica que $\mu(i) P_i^s \mu'(i)$ para todo $i \in T$. Además por la definición de preferencias egocéntricas se tendría que $\mu P_i \mu'$ para todo $i \in T$, pero esto contradeciría 1. Como los miembros de $M \setminus T$ no cambian de pareja tenemos que $\mu'(i) I_i^s \mu(i)$ para todo $i \in M \cup W \setminus T$. Juntando ambos resultados concluimos que μ no es Pareto eficiente para M cuando las preferencias son R^s. También sabemos que $TTC_M(R^s) = TTC_M^e(R) = \mu$ por lo que el mecanismo TTC_M no es Pareto eficiente para M cuando las preferencias son R^s. Por la Proposición 3 de Abdulkadiroğlu y Sönmez (2003) sabemos que TTC_M es un mecanismo Pareto eficiente para M, lo cual es una contradicción.

Mostraremos que $TTC_M^e : \mathcal{R}^{ego} \to \mathcal{A}(M, W)$ es individualmente racional para M. Supongamos que esto no es cierto. En dicho caso, existe un perfil de preferencias $R \in \mathcal{R}^{ego}$ y un $i \in M$ en donde i bloquea a $TTC_M^e(R) = \mu$. Por la definición de bloqueo tenemos que $\mu'(i) \neq \mu(i)$ y $\mu' P_i \mu$ para todo $\mu' \in \Theta^{\{i\},f}$ donde $f(i) = i$. Dado lo anterior, tenemos que $i P_i^s \mu(i)$ en donde P_i^s es la preferencia estándar inducida de la preferencia egocéntrica R_i. Como $\mu = TTC_M^e(R)$, tenemos que en la etapa en que salió i, este apuntó a una persona en W que era menos preferida que apuntar a si mismo, lo cual es una contradicción con las etapas del algoritmo.

Por último mostraremos que $TTC_M^e : \mathcal{R}^{ego} \to \mathcal{A}(M, W)$ es *strategy-proof*. Supongamos que esto no es cierto. Esto es, existe un problema de *marriage market* con externalidades cuyo perfil de preferencias es $R \in \mathcal{R}^{ego}$ y en donde existe un $i \in M$, en el que $TTC_M^e(\tilde{R}_i, R_{-i}) P_i TTC_M^e(R)$. Sea R^s el perfil de preferencias estándar inducido de R y R_i^s la preferencia estándar inducida de i. Como las preferencias son egocéntricas existen dos casos posibles:

- El primero corresponde a $TTC_M^e(\tilde{R}_i, R_{-i})(i) P_i^s TTC_M^e(R)(i)$. Como $TTC_M^e(R) = TTC_M(R^s)$, esto implicaría que TTC_M con preferencias estándar no es *strategy-proof* para M, lo cual es una contradicción con la Proposición 4 de Abdulkadiroğlu y Sönmez (2003).

- El segundo corresponde a $TTC_M^e(\tilde{R}_i, R_{-i})(i) = TTC_M^e(R)(i)$, pero que i mejore producto de que alguien cambió de pareja. Como $TTC_M^e(R) = TTC_M(R^s)$, lo anterior implica que TTC_M^e y TTC_M no son mecanismos *non-bossy*. Por Pápai (2000) sabemos que TTC_M es *non-bossy*, lo cual es una contradicción.[23]

$\square$

[23] Pápai (2000) muestra que una regla, en el contexto de housing market, es Pareto eficiente, *strategy-proof*, *non-bossy* y *reallocation-proofness* si y solamente si es una *hierarchical exchange rule*, donde en este caso tenemos que TTC_M corresponde a un caso particular de una *hierarchical exchange rule*.

Teorema 3

Demostración. Partiremos mostrando que si en un problema de *marriage market* con externalidades, el perfil de preferencias es $R \in \mathcal{R}_M^{ego} \times \mathcal{R}_W^{ego-w}$, entonces ni un miembro en M puede reportar preferencias falsas y mejorar cuando se implementa el algoritmo AD_M^e. Supongamos que esto no es cierto. Esto es, existe $R \in \mathcal{R}_M^{ego} \times \mathcal{R}_W^{ego-w}$ y un $i \in M$ que cumple con $AD^e(\tilde{R}_i, R_{-i})P_i AD^e(R)$. Sea $R^s = (R_i)_{i \in M \cup W}$ el perfil de preferencias estándar inducido de R y R_i^s la preferencia estándar inducida del individuo i. Como las preferencias son egocéntricas existen dos casos posibles:

- El primer caso corresponde a que el individuo i mejora a través de cambiar su pareja. Esto es, $AD_M^e(\tilde{R}_i, R_{-i})(i)P_i^s AD_M^e(R)(i)$. Como $AD_M^e(R) = AD_M(R^s)$, esto implica que algún i mejoró su situación a través de reportar preferencias falsas. Por lo tanto, AD_M con preferencias estándar no es *strategy-proof* para M. Esto último es una contradicción con el Teorema 9 de Dubins y Freedman (1981).

- El segundo caso corresponde a que el individuo i mantiene su pareja, pero mejora a través de cambiar las parejas del resto. En ese caso tenemos que $AD^e(\tilde{R}_i, R_{-i})(i) = AD^e(R)(i)$ y que $AD^e(\tilde{R}_i, R_{-i})(j) \neq AD^e(R)(j)$ para un $j \neq i$. Como $AD_M^e(R) = AD_M(R^s)$, lo anterior implica que AD_M^e y AD_M^e no son mecanismos *non-bossy*. Esto último contradice Ergin (2002), el cual muestra que AD_M es *non-bossy* para M en caso de que no hayan ciclos en el perfil de preferencias de W.[24]

Mostremos ahora que si en un problema de *marriage market* con externalidades, el perfil de preferencias es $R \in \mathcal{R}_M^{ego} \times \mathcal{R}_W^{ego-w}$, entonces el matching $AD_M^e(R)$ es Pareto eficiente para M. Supongamos que esto no es cierto. Entonces existe un perfil $R \in \mathcal{R}_M^{ego} \times \mathcal{R}_W^{ego-w}$ en donde $AD_M^e(R) = \mu$ no es Pareto eficiente para M. En ese caso existe un matching μ' que cumple con:

1. $\mu' R_i \mu$ para todo $i \in M$.

2. $\mu' P_i \mu$ para un $i \in M$.

Sea $T = \{i \in M : \mu'(i) \neq \mu(i)\}$. Aprendemos que $T \neq \emptyset$, caso contrario se contradeciría 2. Sea $R^s = (R_i)_{i \in M \cup W}$ el perfil de preferencias estándar inducido de R y R_i^s la preferencia estándar inducida del individuo i. Como las preferencias son egocéntricas tenemos que las preferencias estándar inducidas cumplen con $\mu'(i)P_i^s\mu(i)$ para todo $i \in T$. Caso contrario se tendría que $\mu(i)R_i^s\mu'(i)$, pero como $\mu(i) \neq \mu'(i)$, implica que $\mu(i)P_i^s\mu'(i)$ para todo $i \in T$. Además por la definición de preferencias egocéntricas se tendría que $\mu P_i \mu'$ para todo $i \in T$, pero esto contradeciría 1. Como los miembros de $M \setminus T$ no cambian de pareja, tenemos que $\mu'(i)I_i^s\mu(i)$ para todo $i \in M \setminus T$. Juntando ambos resultados concluimos que μ' no es Pareto eficiente para M cuando las preferencias son R^s. Sabemos que $AD_M(R^s) = AD_M^e(R)$, por lo que AD_M no es Pareto eficiente para M cuando el perfil de preferencias es R^s. Esto último contradice Teorema 1 de Ergin (2002), el cual dice que en ausencia de ciclos en el

[24]Recordar que en el caso sin externalidades diremos que el perfil de preferencias de W no tiene ciclos si el perfil de preferencias R_W^s es la preferencia estándar inducida de algún $R_W \in \mathcal{R}_W^{ego-w}$.

perfil de preferencias de W, el matching $AD_M(R)$ es Pareto eficiente para M.

Ahora mostraremos que si AD^e_M es *strategy-proof* para M, entonces $\mathcal{R}^{ego-w}_W$ es el dominio maximal en $\mathcal{R}^{ego}_W$ donde sin hacer hipótesis sobre $\mathcal{R}^{ego}_M$ se cumple dicha propiedad. Supongamos que esto no es cierto. Esto es, existe un perfil de preferencias egocéntricas de W, $R_W \in \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w}_W$ en donde para todo $R_M \in \mathcal{R}^{ego}_M$ no hay miembros de M que tengan incentivos a reportar preferencias falsas y mejorar cuando se implementa AD^e_M. Si $R_W \in \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w}_W$, entonces existen $w_1, w_2 \in W$ y $m_1, m_2, m_3 \in M$ donde las preferencias estándar inducidas de las preferencias egocéntricas de w_1 y w_2 satisfacen la siguiente propiedad:

$$m_1 P^s_{w_1} m_2 P^s_{w_1} m_3 P^s_{w_2} m_1.$$

Asumamos que para todo miembro en $M \setminus \{m_1, m_2, m_3\}$ nadie es aceptable y que las preferencias estándar inducidas de las preferencias egocéntricas de m_1, m_2 y m_3 son:

m_1	m_2	m_3
w_2	m_2	w_1
w_1		w_2
m_1		m_3

Asumamos que $\mu' P_{m_2} \mu$, donde μ' y μ son:

$$\mu = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_2 & m_2 & w_1 \end{bmatrix}, \qquad \mu' = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & m_2 & w_2 \end{bmatrix}.$$

Como tanto m_1 como m_3 son aceptables para w_1 y w_2, el resultado de aplicar AD^e_M es μ. Sin embargo, vemos que m_2 podría reportar una preferencia egocéntrica cualquiera, en donde la preferencia estándar inducida cumpla con $\tilde{R}^s_{m_2} : w_1$. En este nuevo caso, note que al momento en que m_3 haga su propuesta a w_1, el individuo m_2 tiene mayor prioridad por lo que logra estar momentáneamente emparejado con w_1. Por su parte m_1 se empareja momentáneamente con w_2. Sabemos que se cumple que $m_3 P^s_{w_2} m_1$, por lo que cuando m_3 le haga la propuesta a w_2, desplazará a m_1. Finalmente sabemos que se cumple $m_1 P^s_{w_1} m_2$, por lo que m_1 desplaza a m_2 y este queda finalmente sin pareja. Dado lo anterior, μ' es el resultado de aplicar AD^e_M cuando m_2 reporta $\tilde{R}_{m_2}$. Por lo tanto AD^e_M no es *strategy-proof* para los miembros de M, ya que $AD^e_M(\tilde{R}_{m_2}, R_{-m_2}) P_{m_2} AD^e_M(R)$.

Mostraremos que si AD^e_M es Pareto eficiente para para M, entonces $\mathcal{R}^{ego-w}_W$ es el dominio maximal en $\mathcal{R}^{ego}_W$ donde sin hacer hipótesis sobre $\mathcal{R}^{ego}_M$ se cumple dicha propiedad. Supongamos que esto no es cierto. Esto es, existe un perfil de preferencias egocéntricas $R_W \in \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w}_W$ en donde para todo $R_M \in \mathcal{R}^{ego}_M$, el matching $AD^e(R)$ es Pareto eficiente para los miembros de M. Si $R_W \in \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w}_W$, entonces existen $w_1, w_2 \in W$ y $m_1, m_2, m_3 \in M$ donde las preferencias estándar inducidas de las preferencias egocéntricas de w_1 y w_2 cumplen la siguiente propiedad:

$$m_1 P^s_{w_1} m_2 P^s_{w_1} m_3 P^s_{w_2} m_1.$$

Asumamos que para todo miembro en $M \setminus \{m_1, m_2, m_3\}$ nadie es aceptable y que las preferencias estándar inducidas de las preferencias egocéntricas de m_1, m_2 y m_3 son:

m_1	m_2	m_3
w_2	w_1	w_1
w_1	m_2	w_2
m_1		m_3

Asumamos que $\mu I_{m_2} \mu'$, donde μ y μ' son:

$$\mu = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_2 & m_2 & w_1 \end{bmatrix}, \qquad \mu' = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & m_2 & w_2 \end{bmatrix}.$$

En este caso el resultado de aplicar AD^e_M es μ'. Podemos notar que μ domina en sentido de Pareto a μ', lo cual es una contradicción con lo que supusimos en un inicio.

Finalmente mostraremos lo relacionado a los incentivos de grupos en M (*group strategy-proof* para M). Si AD^e_M es *group strategy-proof* para M, entonces $\mathcal{R}^{ego-w}_W$ es el dominio maximal en $\mathcal{R}^{ego}_W$ donde sin hacer hipótesis sobre $\mathcal{R}^{ego}_M$ se cumple dicha propiedad. Supongamos que esto no es cierto. Esto es, existe un perfil de preferencias egocéntricas $R_W \in \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w}_W$ en donde para todo $R_M \in \mathcal{R}^{ego}_M$ no hay grupos de miembros de M que tengan incentivos a reportar preferencias falsas y mejorar cuando se implementa AD^e_M. Vimos que si $R_W \in \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w}_W$, entonces existe un perfil $R_M \in \mathcal{R}^{ego}_M$ en donde hay un miembro en M que puede reportar preferencias falsas y mejorar cuando se implementa AD^e_M. Como existe un perfil $R \in \mathcal{R}^{ego}_M \times \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w}_W$ en donde AD^e_M no es *strategy-proof* para M, podemos concluir que para ese mismo perfil, una coalición de un individuo puede mejorar al reportar preferencias falsas, por lo que AD^e_M no sería *group strategy-proof* para M.

Veamos si $R \in \mathcal{R}^{ego}_M \times \mathcal{R}^{ego-w}_W$, entonces AD^e_M es *group strategy-proof* para M. Supongamos que esto no es cierto. Entonces existe un perfil $R \in \mathcal{R}^{ego}_M \times \mathcal{R}^{ego-w}_W$ y un grupo $S \subseteq M$ donde se cumple:

1. $AD^e_M(\tilde{R}_S, R_{-S}) R_i AD^e_M(R)$ para todo $i \in S$.
2. $AD^e_M(\tilde{R}_S, R_{-S}) P_i AD^e_M(R)$ para algún $i \in S$.

Aprendemos que $AD^e_M(\tilde{R}_S, R_{-S}) \neq AD^e_M(R)$, caso contrario no se cumpliría 2. Esto implica que $T = \{i \in M : AD^e_M(\tilde{R}_S, R_{-S})(i) \neq AD^e_M(R)(i)\} \neq \emptyset$. Sea $R^s = (R_i)_{i \in M \cup W}$ el perfil de preferencias estándar inducido de R. Analizaremos dos casos posibles.

- $T \cap S = \emptyset$. Esto quiere decir que los miembros de la coalición mejoraron producto del cambio de la pareja de alguien de afuera, pero los integrantes de la coalición S mantuvieron sus parejas. Como $AD^e_M(R) = AD_M(R^s)$, implica que AD^e_M y AD^e_M no son mecanismos *group non-bossy* para M. Usando solamente las preferencias estándar, sabemos por el Teorema 1 de Ergin (2002) que AD_M es *group strategy-proof* cuando no hay ciclos en el perfil de preferencias de W. Por el Teorema 2 de Afacan (2011) sabemos que si un mecanismo es *group strategy-proof* entonces es *group non-bossy*. Esto contradice lo anterior, ya que en ausencia de ciclos en el perfil de preferencias de W, AD_M es un mecanismo *group non-bossy* para M.

- $T \cap S \neq \emptyset$. Esto quiere decir que hay miembros de la coalición S que cambiaron sus parejas. Notemos que se tiene $AD^e_M(\tilde{R}_S, R_{-S})(i) P^s_i AD^e_M(R)(i)$ para todo $i \in T \cap S$. Caso contrario se tendría que $AD^e_M(R)(i') P^s_{i'} AD^e_M(\tilde{R}_S, R_{-S})(i')$ para al menos un $i' \in T \cap S$. En ese caso por la definición de preferencia egocéntrica, se tendría que para i' se cumple $AD^e_M(R) P_{i'} AD^e_M(\tilde{R}_S, R_{-S})$. Esto último no cumpliría con 1. Sabemos también que $AD^e_M(R)(i) = AD^e_M(\tilde{R}_S, R_{-S})(i)$ para todo $i \in S \setminus T$. Como $AD^e_M(R) = AD_M(R^s)$, lo anterior implica que AD_M no es un mecanismos *group strategy-proof* para M, ya que se tiene que $AD^e_M(\tilde{R}_S, R_{-S})(i) P^s_i AD^e_M(R)(i)$ para todo $i \in T \cap S$ y $AD^e_M(\tilde{R}_S, R_{-S})(i) I^s_i AD^e_M(R)(i)$ para todo $i \in S \setminus T$. Lo anterior es una contradicción con el Teorema 1 de Ergin (2002), el cual dice que AD_M es un mecanismo *group strategy-proof* para M en ausencia de ciclos en el perfil de preferencias de W.

□

Proposición 3.4

Demostración. Primero probaremos que si $R \in \mathcal{R}^{ego}_M \times \mathcal{R}^{ego-w-f}_W$, entonces TTC^e_M es un mecanismo estable. Suponga que existe un $R \in \mathcal{R}^{ego}_M \times \mathcal{R}^{ego-w-f}_W$ en donde $TTC^e_M(R)$ no es matching estable. Sea R^s el perfil de preferencias estándar inducido. Por el Teorema 1 de Kesten (2006) sabemos que $AD_M(R^s) = TTC_M(R^s)$. Como $TTC^e_M(R) = TTC_M(R^s)$, tenemos que $AD_M(R^s)$ no es un matching estable, lo cual es una contradicción con el Teorema 1 de Gale y Shapley (1962).

Ahora probaremos que si TTC^e_M es estable, entonces $\mathcal{R}^{ego-w-f}_W$ es el dominio maximal en $\mathcal{R}^{ego}_W$ donde sin hacer hipótesis sobre $\mathcal{R}^{ego}_M$ se cumple dicha propiedad. Supongamos que esto no es cierto. Esto es, existe un perfil de preferencias egocéntricas $R_W \in \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w-f}_W$ en donde para todo $R_M \in \mathcal{R}^{ego}_M$, el matching $TTC^e_M(R)$ es estable. Si $R_W \in \mathcal{R}^{ego}_W \setminus \mathcal{R}^{ego-w-f}_W$, entonces existen $w_1, w_2 \in W$ y $m_1, m_2, m_3 \in M$ donde las preferencias estándar inducidas a partir de las preferencias egocéntricas de w_1 y w_2 cumplen con $m_1 P^s_{w_1} m_2 P^s_{w_1} m_3$ y al menos una de las siguientes propiedades:

- $m_3 P^s_{w_2} m_1$.
- $m_3 P^s_{w_2} m_2$.

Asuma que para todo $M \setminus \{m_1, m_2, m_3\}$ nadie es aceptable y que las preferencias estándar inducidas de las preferencias egocéntricas de m_1, m_2 y m_3 son:

m_1	m_2	m_3
w_2	w_1	w_1
m_1	m_2	m_3

Vemos que tanto para la condición $m_3 P^s_{w_2} m_1$ como para la condición $m_3 P^s_{w_2} m_2$, el resultado de aplicar TTC^e_M es el siguiente:

$$\mu = \begin{bmatrix} m_1 & m_2 & m_3 \\ w_2 & m_2 & w_1 \end{bmatrix}.$$

Vemos que μ no es estable ya que puede ser bloqueado por el par $\{m_2, w_1\}$. Esto debido a que se cumple que $\eta P_i \mu$, para $i \in \{m_2, w_1\}$ y para todo $\eta \in \Theta^{\{m_2,w_1\},f}(\mu)$, donde $f(m_2) = w_2$, lo cual es una contradicción de que $TTC^e_M(R)$ es un matching estable. $\square$

<u>*Teorema 5*</u>

Demostración. Sea $S_1, \dots, S_n$ las n coaliciones disjuntas de $M \cup W$. Además f_τ, con $\tau \in \{1, 2, ..., n\}$, corresponde al acuerdo que es top o dominante para los miembros S_τ. Como cada acuerdo f_τ es una función biyectiva $f_\tau : S_\tau \to S_\tau$ y sabemos que se cumple que $\bigcap_{\tau=1}^n S_\tau = \emptyset$, entonces el matching resultante de $\bigcap_{\tau=1}^n \mathcal{A}^{S_\tau, f_\tau}$ es único y al cual llamaremos μ.

Notemos que, independiente de si las preferencias son top o dominantes, todo matching $\mu' \in \mathcal{A}(M, W) \backslash \{\mu\}$ es bloqueado por al menos por una coalición S_τ con $\tau \in \{1, 2, ..., n\}$. Si es que $\mu' \neq \mu$, significa que $T' = \{i \in M \cup W : \mu(i) \neq \mu'(i)\}$ es diferente de vacío. Considere a $i' \in T'$ como la persona con el menor índice entre los miembros de T'. Sea j' el índice de i' (si hay más de una persona en $S_{j'}$ elegir alguno cualquiera). Notemos que como en $S_{j'}$ están las personas con menor índice en T', tenemos que $\mu' \in \mathcal{A}(M, W) \setminus \mathcal{A}^{S_j, f_j} \cup \mathcal{D}_i$ para todo $i \in S_{j'}$. Sabemos además que las preferencias del grupo $S_{j'}$ son top o dominantes, por lo que tenemos que $\eta P_i \mu'$ para todo $\eta \in \Theta^{S_{j'}, f_{j'}}$ y para todo $i \in S_{j'}$. Por lo tanto dicho grupo bloquearía a μ'. Esto último es equivalente a que la coalición $S_{j'}$ bloquee μ' a través del acuerdo $f_{j'}$.

Ahora mostraremos que μ está en el $\Theta-core$. Supongamos que μ es bloqueado por una coalición T, través de un acuerdo g. Entonces se tiene que:

1. $\mu(i) \neq g(i)$ para algún $i \in T$.
2. $\eta R_i \mu$ para todo $\eta \in \Theta^{T,g}(\mu)$ y para todo $i \in T$.
3. Para todo $\eta \in \Theta^{T,g}(\mu)$ existe un $i \in T$ que $\eta P_i \mu$.

Por 1 aprendemos que $T' = \{i \in T : \mu(i) \neq g(i)\}$ es diferente de vacío. Sea $i' \in S_j$ alguien en T' donde i' es la persona con menor índice entre los miembros de T' (si hay más de una persona de S_j en T' elegir alguno cualquiera). Podemos notar que para todo $\eta \in \Theta^{T,g}(\mu)$ se tiene que $\eta(i') \in S_{j'}$ con $j' \geq j$. Lo anterior implica que $\eta \in \mathcal{A}(M, W) \setminus \mathcal{A}^{S_j, f_j} \cup D_{i'}$ para todo $\eta \in \Theta^{T,g}(\mu)$. Sabemos también que las preferencias del grupo S_j son top o dominantes, por lo que se cumple que $\mu P_{i'} \mu'$ para todo $\mu' \in \mathcal{A}(M, W) \setminus \mathcal{A}^{S_j, f_j} \cup D_{i'}$. Como caso particular de esto último, se tiene que $\mu P_{i'} \eta$ para todo $\eta \in \Theta^{T,g}(\mu)$. Podemos ver que lo anterior es una contradicción con el punto 2, por lo que la coalición T no pudo haber bloqueado a μ. $\square$

Apéndice B

Cuadro 1: Tabla de símbolos

Símbolo	Significado	Página
$\mathcal{A}(M,W)$	Conjunto de todos los matchings entre los miembros de M y W	4
R_i	Preferencia del agente i	4
$\mathcal{R}_i$	Conjunto de preferencias del agente i	5
$T \subseteq M \cup W$	Coalición de miembros de $M \cup W$	5
R_T	Perfil de preferencias de una coalición T	5
R	Perfil de preferencias de $M \cup W$	5
$\mathcal{R}_T$	Conjunto de perfiles de preferencias de una coalición T	5
$\mathcal{R}$	Conjunto de perfiles de preferencias	5
$\mathcal{A}(T)$	Matchings en $\mathcal{A}(M \cap T, W \cap T)$	5
$\mathcal{A}^{T,f}$	Matchings en $\mathcal{A}(M,W)$ en los cuales la coalición T siguen el acuerdo f	5
$\Theta^{T,f}(\mu)$	Creencias de T sobre posibles reacciones del resto cuando ellos siguen el acuerdo f	5
$\mathcal{C}(R)$	Conjunto de matchings en el $\Theta-$*core* para el perfil de preferencias R	5
$\mathcal{S}(R)$	Conjunto de matchings estables para el perfil de preferencias R	6
$\mathcal{R}_T^{ego}$	Conjunto de perfiles de preferencias egocéntricas de T	8
$\mathcal{R}^{ego}$	Conjunto de perfiles de preferencias egocéntricas	8
R_i^s	Preferencia estándar inducida por $R_i \in \mathcal{R}_i^{ego}$	8
R_T^s	Perfil de preferencias estándar de T	8
R^s	Perfil de preferencias estándar de $M \cup W$	8
$\mathcal{R}_T^s$	Conjunto de perfiles de preferencias estándar de T	8
$\mathcal{R}^s$	Conjunto de perfiles de preferencias estándar	8
AD_M	Algoritmo de aceptación diferida cuando M hace las propuestas	8
TTC_M	Algoritmo *top trading cycle* cuando se prioriza a M	8
AD_M^e	Algoritmo de aceptación diferida con preferencias egocéntricas cuando M propone	9
TTC_M^e	Algoritmo *top trading cycle* con preferencias egocéntricas cuando se prioriza a M	9
$\mathcal{R}_W^{ego-w}$	Conjunto de perfiles de preferencias egocéntricas de W que son w-acíclicas	17
$\mathcal{R}_W^{ego-w-f}$	Conjunto de perfiles de preferencias egocéntricas de W que son w-acíclicas fuerte	19
R_i^*	Preferencia aguas abajo del agente i	19
$\mathcal{R}_i^*$	Conjunto de preferencias aguas abajo del agente i	20
$\mathcal{R}_T^*$	Conjunto de perfiles de preferencias aguas abajo de una coalición T	20
AD_M^*	Algoritmo de aceptación diferida con creencias extremamente prudentes cuando M propone	20
$\mathcal{R}^t$	Conjunto de perfiles de preferencias top	23
$\mathcal{D}_i$	Conjunto de matchings en donde el agente i está emparejado con alguien de un índice menor	23
$\mathcal{R}^d$	Conjunto de perfiles de preferencias dominantes	23
$\mathcal{B}^H$	*Bargaining set* a la Hafalir	26
$\mathcal{B}$	*Bargaining set*	26