



Jornadas de Investigación en Filosofía

Departamento de Filosofía. Facultad
de Humanidades y Ciencias de la Educación.
Universidad Nacional de La Plata

Decisiones sociales basadas en argumentación. Algunos resultados pesimistas

Gustavo A. Bodanza (UNacionalSur - CONICET)

Introducción

Es natural pensar en la argumentación como una de los ingredientes principales de cualquier forma de decisión social racional. Idealmente, los individuos deben debatir y las decisiones del grupo deben estar basadas en los argumentos que el grupo como un todo considera mejor justificadas. El problema que nos ocupará es el de cómo el grupo debe proceder para hallar tales decisiones. Mostraremos que distintos mecanismos, todos en apariencia razonables, pueden llevar a resultados opuestos (de modo análogo a cómo surge la paradoja doctrinal en la agregación de juicios). Por otra parte, en la medida en que los mecanismos de agregación son justos, en el sentido que cumplen con los requisitos de racionalidad planteados por Arrow en la Teoría de la Elección Social, las decisiones aceptables sólo podrán ser alcanzadas mediante acuerdos unánimes. En este sentido, los resultados que mostraremos parecerán desalentadores.

En este trabajo combinamos el modelo de marcos argumentativos de Dung (1995) con la Teoría de la Elección Social. El primero nos permitirá modelar situaciones en las que, a partir de un conjunto de *argumentos* (tomados en abstracto) que interactúan mediante una relación de *derrota*, se puede obtener un subconjunto de argumentos “ganadores”. Esto servirá para modelar tanto las elecciones que hace individualmente cada agente de los argumentos mediante criterios de derrota personales, como la elección que hará el grupo como un todo mediante criterios colectivos de derrota. La Teoría de la Elección Social, por su parte, nos permitirá analizar la aceptabilidad de los mecanismos de agregación empleados en base a las propiedades arrobianas.

Ejemplo motivador

Supongamos que un equipo de tres médicos (M_1 , M_2 y M_3) debe decidir qué terapia aplicar a un paciente que presenta determinados síntomas. En la deliberación surgen

principalmente los siguientes tres argumentos:

A: "Los síntomas x , y y z sugieren la presencia de la enfermedad e_1 , por lo tanto debemos aplicar la terapia t_1 ."

B: "Los síntomas x , w y z sugieren la presencia de la enfermedad e_2 , por lo tanto debemos aplicar la terapia t_2 ."

C: "Los síntomas x y z sugieren la presencia de la enfermedad e_3 , por lo tanto debemos aplicar la terapia t_3 ."

Ahora supongamos que cada uno de los médicos hace su propia ponderación de los argumentos, resultando lo siguiente:

M1 considera inconveniente aplicar conjuntamente las terapias t_1 y t_2 o las terapias t_2 y t_3 ; además cree que el argumento B es más específico que el C, por lo que concluye que **B derrota a C**; y cree que en el caso en cuestión el síntoma y se manifiesta más claramente que el w , concluyendo que **A derrota a B**.

M2 considera inconveniente aplicar conjuntamente la terapia t_2 con t_1 o con t_3 ; además cree que el síntoma w no se manifiesta claramente, por lo tanto **C derrota a B**; y cree que en el caso en cuestión el síntoma w se manifiesta más claramente que y , concluyendo que **B derrota a A**.

M3 considera inconveniente aplicar la terapia t_2 conjuntamente con t_1 o con t_3 ; además cree que el síntoma w se manifiesta más claramente que y ; también considera que el argumento B es más específico que el C, por lo que concluye que **B derrota a A** y **B derrota a C**.

Vamos a considerar ahora dos procedimientos distintos mediante los cuales el grupo puede tomar la decisión:

Procedimiento P1: *sobre cada par de argumentos, cada agente emitirá su juicio acerca de si éstos se derrotan o no y de qué modo. Luego se obtiene una relación (criterio) de derrota colectiva agregando las relaciones (criterios) individuales. Finalmente, se eligen los argumentos ganadores en base a esta relación colectiva.*

Procedimiento P2: *cada agente vota por los argumentos que considera ganadores de acuerdo a su propia relación (criterio) de derrota. Luego, los votos de todos los agentes se agregan en un único conjunto de argumentos que serán los socialmente aceptados.*

Se puede argumentar que el procedimiento P1 busca que la decisión social se base en las razones de los agentes, en tanto no pondera simplemente sus decisiones últimas sino los criterios que los llevan a hacer sus elecciones. Por su parte, el procedimiento P2

parece ser el más corrientemente utilizado en las decisiones sociales, esto es, los individuos simplemente expresan su voto y éstos luego se agregan. Sin embargo, determinar cuál de los dos procedimientos resulta más racional es una cuestión que amerita un análisis filosófico profundo que no estamos en condiciones de ofrecer aquí. Lo que nos vamos a preguntar, en cambio, es si ellos arrojarán siempre los mismos resultados permitiendo que se apliquen indistintamente (por ejemplo, siguiendo el más sencillo), o si arrojarán el mismo resultado bajo ciertas condiciones razonables, o si, en el peor de los casos, no hay condiciones razonables bajo las cuales arrojen el mismo resultado.

Veamos que ocurre en nuestro ejemplo siguiendo cada uno de los procedimientos y suponiendo que la implementación de los mismos se realiza mediante la regla de *mayoría absoluta* (i.e., se imponen las opciones que sean elegidas por más de la mitad de los agentes). Siguiendo el procedimiento **P1**, entonces, tendremos la siguiente situación:

1. Se elige el criterio de derrota:

B derrota a A (por los votos de M2 y M3)

B derrota a C (por los votos de M1 y M2)

A y C no se derrotan entre sí.

2. Se eligen los argumentos ganadores de acuerdo a la relación de derrota obtenida en el paso anterior¹:

Conjunto de argumentos elegidos socialmente: $E_{P1} = \{B\}$

En consecuencia, la decisión del grupo a través del procedimiento P1 implementado mediante la regla de mayoría absoluta será aplicar solamente la terapia $t2$, en tanto el único argumento que se impone es B.

Veamos ahora qué ocurre si se sigue el procedimiento **P2**, implementado también mediante la regla de mayoría absoluta:

1. Cada agente elige argumentos de acuerdo a su propio criterio de derrota²:

Conjunto de argumentos elegidos por M1: {A, C}

Conjunto de argumentos elegidos por M2: {A, C}

Conjunto de argumentos elegidos por M3: {B}

¹ De aquí en adelante consideraremos que los argumentos son elegidos siguiendo la extensión *grounded* del marco argumentativo. Ver el Apéndice para los detalles técnicos.

² Aquí también se obtiene la extensión *grounded* de cada marco argumentativo individual.

2. Los votos de los agentes se agregan en un único conjunto que expresa las elecciones colectivas:

Conjunto de argumentos elegidos socialmente: $E_{P_2} = \{A, C\}$

Por lo tanto, la decisión del grupo a través del procedimiento P2 implementado mediante la regla de mayoría absoluta será aplicar a la vez las terapias t_1 y t_3 , en tanto se imponen sendos argumentos A y C.

Esto muestra que los procedimientos P1 y P2 no coinciden en las decisiones resultantes en general (i.e. $E_{P_1} \neq E_{P_2}$). La siguiente cuestión, entonces, será imponer condiciones razonables de acuerdo a los fines de una elección democrática que sean suficientes para garantizar la coincidencia de ambos procedimientos en las decisiones resultantes.

Posibilidad de coincidencia restringida usando el mecanismo de mayoría absoluta

Hemos visto que los procedimientos P1 y P2 no coinciden en general cuando son implementados mediante el mecanismo de mayoría absoluta. Puesto que este mecanismo es de usual aplicación en las decisiones democráticas, es importante saber si existen condiciones democráticamente adecuadas para que su aplicación indistinta en los procedimientos P1 y P2 conduzca a los mismos resultados. En particular, podría pensarse en consensos mínimos que fueran suficientes para ese propósito.

En Bodanza y Auday (2009) se proponen dos tipos de consenso para lograr la coincidencia de P1 y P2 mediante mayoría absoluta. Uno es más bien débil: para cada par de argumentos A y B, los individuos deben estar de acuerdo en si éstos están en conflicto o no (o sea, si algún individuo cree que A y B no se derrotan en ningún sentido, entonces ningún individuo cree que se derrotan). El otro exige algo más: debe haber al menos dos argumentos en conflicto tales que todos los individuos opinan que se derrotan de la misma manera. El problema es que aún con estos consensos la coincidencia de P1 y P2 no se daría a menos que los individuos sean suficientemente racionales como para no proponer derrotas cíclicas, que haya un número impar de individuos para que no se den empates, y que no haya más de tres argumentos en disputa. Más formalmente:

Resultado 1³ (Bodanza y Auday, 2009). Si se cumplen las condiciones:

1. *regla de mayoría absoluta en P1 y P2,*
2. *no más de tres argumentos*
3. *número impar de individuos*

3 Para la prueba de éste y los siguientes resultados se remite a los correspondientes trabajos citados.

4. \otimes_i es acíclica para cada individuo i (donde ' \otimes_i ' simboliza la relación de derrota del individuo i)
5. para cada par de argumentos, hay unanimidad respecto de si hay conflicto o no entre ellos
6. existe un par de argumentos A y B tales que o bien " i ($A \otimes_i B$) o bien " i ($B \otimes_i A$) entonces $E_{P1} = E_{P2}$.

En nuestra opinión, este resultado muestra que los consensos mínimos que hacen posible la coincidencia junto con las restricciones sobre los números de individuos y argumentos deben ser demasiado restrictivos cuando se quiere aplicar la regla de mayoría absoluta.

Pero ¿qué ocurre con otros mecanismos de votación? En ciertas situaciones de elección social se aplican razonablemente otras reglas. Por ejemplo, en la Argentina es común que cuando los jóvenes organizan fiestas el menú de comidas y bebidas que se consumirán se decide distribuyendo los derechos a elección de modo que las mujeres eligen (y se encargan de llevar) la comida y los hombres eligen (y se encargan de llevar) las bebidas. Como resultado, todos comen lo que eligen las mujeres y beben lo que eligen los hombres. Dicho en términos más técnicos, las mujeres forman un *conjunto decisivo* respecto de la comida y los hombres forman un *conjunto decisivo* respecto de las bebidas. Formalmente, un conjunto de individuos es *decisivo* si cualquier alternativa que es elegida por todos los miembros del conjunto se impone socialmente. Puede ocurrir también que cada individuo se encargue de elegir un ítem (de comida si es mujer y de bebida si es hombre), en cuyo caso cada individuo conforma por sí solo un conjunto decisivo unimembre. Está claro que la primera variante de la regla de votación puede ser caracterizada por la clase de los conjuntos decisivos conformados por mujeres y hombres respectivamente, mientras la segunda variante se puede caracterizar por la clase de todos los conjuntos que contienen a uno y solo un individuo.

En Bodanza y Auday (2009) se encuentran relaciones entre $P1$ y $P2$ cuando son implementados mediante reglas de votación que, en general, pueden caracterizarse por una clase de conjuntos decisivos. Para esto se relajan las restricciones sobre los números de argumentos y de individuos y se ve qué tipos de consensos son suficientes para llegar a una similitud entre $P1$ y $P2$. El resultado es que puede hallarse una relación entre ambos procesos requiriendo solamente un acuerdo respecto de qué argumentos están en conflicto y unanimidad dentro de cada conjunto. Los mecanismos de agregación en cada procedimiento se pueden describir como sigue:

- Relación de derrota agregada \otimes_s : $A \otimes_s B$ si y sólo si existe un conjunto decisivo D tal que para todo individuo $i \in D$, $A \otimes_i B$ (donde \otimes_i es la relación de derrota sostenida por i).

E_{P_1} será la extensión *grounded* del marco argumentativo construido con \mathbb{R}_s .

• Conjunto de argumentos socialmente elegidos: $E_{P_2} = \{A: A \hat{I}_{iD} E_i\}$, donde E_i es el conjunto de argumentos elegidos por el individuo i que es miembro del conjunto decisivo D .

Resultado 2 (Bodanza y Auday, 2009). Sea W la clase de todos los conjuntos decisivos de individuos. Si se cumplen las condiciones

1. *acuerdo acerca de los conflictos entre argumentos,*
2. *para cualquier conjunto decisivo $D \hat{I} W$ e individuos $i, j \hat{I} D$,*
 - $\mathbb{R}_i = \mathbb{R}_j$ (acuerdo intra conjuntos decisivos sobre las derrotas) y
 - para todo i, \mathbb{R}_i no tiene ciclos,

entonces $E_{P_1} \hat{I}_{iD} E_i$ (o sea, los argumentos elegidos por P_1 son un subconjunto de los elegidos por P_2)⁴.

Este resultado muestra que los argumentos que un conjunto decisivo de individuos logre imponer mediante P_2 comprenderán a los argumentos que se impongan mediante P_1 , aunque no se dará necesariamente una coincidencia.

Mecanismos de agregación “justos”

En su tesis doctoral en los años '50, Kenneth Arrow propuso una serie de condiciones que hacen razonable una agregación de preferencias individuales (Arrow, 1963). Su resultado negativo muestra que si las preferencias individuales son reflexivas y transitivas entonces no puede haber ningún mecanismo de agregación que dé lugar a una preferencia social reflexiva y transitiva que cumpla con todas sus condiciones. Las condiciones arrowianas resultan razonables también para los procedimientos de agregación P_1 . Sólo que el resultado negativo de Arrow no tiene por qué darse aquí ya que las relaciones de derrota que se agregan no son reflexivas ni transitivas y, en consecuencia, tampoco debemos esperar una relación de derrota agregada reflexiva y transitiva.

Las condiciones de Arrow podemos plantearlas de la siguiente manera:

Optimalidad de Pareto: Si cada individuo decide que A derrota a B , entonces la sociedad decide que A derrota a B .

Asociación Positiva de Valores Individuales: Si I es el conjunto de todos los individuos que votan ' A derrota a B ' y ' A derrota a B ' resulta socialmente decidido, entonces, para

⁴ Estrictamente, lo que dice el resultado original es que $I_{iD} E_i$ es una extensión preferida del marco argumentativo con derrotas \mathbb{R}_s (ver el Apéndice).

cualquier conjunto de individuos J que incluyera a I , si todos los individuos de J votaran 'A derrota a B' entonces 'A derrota a B' resultaría socialmente decidido.

Independencia de las Alternativas Irrelevantes: Si en dos situaciones distintas, todos los individuos eligen exactamente igual con respecto a un par de argumentos A y B, entonces la decisión social respecto de ese par de argumentos debe ser la misma en ambas situaciones.

No dictadura: No existe ningún individuo que logre imponer socialmente su criterio de derrota para todo par de argumentos y para todo conjunto de elecciones individuales.

El resultado que mostraremos ha sido expuesto por Tohmé, Bodanza y Simari (2008) y se basa en resultados previos obtenidos por Brown (1974, 1975) en teoría de la elección social. Brown demuestra que si los votantes ordenan sus preferencias sin que se produzcan ciclos, entonces cualquier mecanismo de agregación que cumpla con las condiciones de arriba producirá una preferencia social también acíclica (lo que elimina serios problemas). Pero lo más interesante, es que la estructura de la clase de conjuntos decisivos determinará la existencia de un pequeño grupo de individuos con poder de veto. La adaptación de este resultado a relaciones de derrota entre argumentos –en lugar de relaciones de preferencia entre alternativas en general- no es trivial. Esto se debe a que las relaciones de derrota no cumplen ninguna de las propiedades usualmente atribuibles a las relaciones de preferencia. Veamos, por caso, la transitividad: si una alternativa x es preferida a una alternativa y y y es preferida a z , entonces es razonable esperar que x sea preferida a z . Sin embargo, si un argumento A derrota a un argumento B y B derrota a un argumento C, de esto no se sigue que A derrote a C (más bien, A parece “defender” a C). Entonces, lo que Tohmé *et al.* (2008) muestran es que, a pesar de estas diferencias en las propiedades de las preferencias y las derrotas entre argumentos, el resultado de Brown se mantiene. O sea,

Resultado 3 (Tohmé et al., 2008). Si un mecanismo de votación sobre criterios de derrota entre argumentos satisface Optimalidad de Pareto, Asociación Positiva de Valores Individuales, Independencia de las Alternativas Irrelevantes y No dictadura entonces va a existir un conjunto de individuos que pertenecen a todo conjunto decisivo.

La justificación de este teorema se da mostrando que si se cumplen las condiciones antecedentes del enunciado, entonces la clase de conjuntos decisivos del mecanismo de votación que se trate formará un *prefiltro propio*, esto es, se darán las siguientes condiciones:

1) $V \setminus W$,

2) si $I \hat{=} W F$ y $I \hat{=} J$, entonces $J \hat{=} W$,

3) $\forall W \exists A \in$,

donde V es el conjunto de todos los votantes y W es la clase de todos los conjuntos decisivos. La condición 3 dice que todos los conjuntos decisivos tienen al menos un individuo en común. Esto significa que siempre habrá un grupo pequeño de individuos que, al pertenecer a todo grupo decisivo, tendrán poder de veto. Ocasionalmente se ha interpretado a este grupo como una “dictadura oculta”; sin embargo parece más apropiado interpretarlo como un *collegium* (ver Brown, 1974). Un *collegium* no es una oligarquía, porque a pesar de pertenecer a todo conjunto decisivo, no necesariamente es decisivo por sí mismo. Entonces, cualquier decisión social debe ser necesariamente apoyada por el *collegium*, pero es suficiente con que el *collegium* rechace una propuesta para que la misma no se imponga socialmente. Esto es lo que ocurre, por ejemplo, en los sistemas parlamentarios como el argentino, donde el Presidente de la Nación conforma el *collegium* ya que pertenece a todo conjunto decisivo para la aprobación de leyes.

Como consecuencia, si se pretende que todo mecanismo justo de agregación de derrotas deba evitar que algunos individuos tengan tal poder de veto, entonces el teorema asegura que no existe ningún mecanismo así. En este sentido, es un resultado de imposibilidad.

Por qué los resultados son pesimistas

Finalmente, mostraremos las consecuencias que se desprenden de los dos últimos resultados expuestos. Recapitulando, el resultado 2 dice que si los miembros de un conjunto decisivo de individuos logran acordar cuáles son las derrotas y éstas son racionales (i.e. acíclicas), entonces entre los argumentos que ellos elegirían individualmente (P2) estarán aquellos que la sociedad aceptaría como consecuencia de una relación de derrota agregada (P1). Sin embargo, el resultado 3 muestra que si la derrota agregada (por P1) se obtiene mediante un mecanismo justo –en sentido arroviado- entonces habrá un conjunto de individuos que pertenecen a todo conjunto decisivo. En consecuencia, si la derrota agregada (por P1) se obtiene mediante un mecanismo justo y se dan los acuerdos expresados en las condiciones del resultado 2, entonces habrá unanimidad total entre todos los miembros de todos los conjuntos decisivos respecto de cuáles son las derrotas; porque si en cada conjunto decisivo hay unanimidad y hay individuos comunes a todos los conjuntos decisivos, entonces es obvio que todos los individuos piensan igual. La siguiente consecuencia es que bajo las condiciones de ambos resultados se recupera la coincidencia entre P1 y P2, pero el precio de esto –y he aquí lo pesimista de la conclusión- es que todos los individuos deben estar absolutamente de acuerdo.

Conclusiones

Hemos comentado, primero, algunos resultados previos que muestran que con la regla de mayoría absoluta no se puede garantizar que los procedimientos P1 y P2 coincidan salvo con fuertes restricciones (resultado 1) y que algunas reglas de votación distintas permiten elegir mediante P2 lo mismo que con P1 y tal vez algo más, siempre que haya unanimidad en todos los conjuntos decisivos (resultado 2). Luego mostramos que en una agregación de derrotas (P1) justa habrá individuos que pertenecen a todo conjunto decisivo (resultado 3). Finalmente, obtuvimos la conclusión que se necesitará unanimidad total si se quiere lograr la coincidencia de los procedimientos siempre que éstos sean justos.

La búsqueda de un escape a este resultado puede llevar a distintas exploraciones, por ejemplo, buscar fallas en extrapolar las condiciones de Arrow para la agregación de preferencias a la agregación de criterios de derrota. También podría optarse por hacer un análisis filosófico en profundidad de los procedimientos P1 y P2 que pondere su adecuación o no a los fines de una democracia deliberativa, de modo que si se observa la inadecuación de alguno de ellos se disuelva el problema de la coincidencia. Otros enfoques del problema pueden llevar a aceptar estos resultados y pensar que la argumentación debe operar a un nivel persuasivo a fin de alcanzar el consenso y obviar el problema de la agregación (este parece ser el punto de vista de Elster⁵).

Referencias

ARROW, K.J. (1963) *Social Choice and Individual Values*. Wiley.

BODANZA, G., AUDAY, M. (2009) "Social Argument Justification: Some Mechanisms and Conditions for Their Coincidence". *Proc. ECSQARU 2009*, Verona, Italia, 95–106.

BROWN, D.J. (1974) "An approximate solution to Arrow's problem". *Journal of Economic Theory* 9, 375—383.

Brown, D.J. (1975) "Aggregation of preferences". *Quarterly Journal of Economics* 89, 456—469.

ELSTER, J. (ed.) (1998) *Deliberative Democracy*. Cambridge University Press.

TOHMÉ, F., BODANZA, G., SIMARI, G. (2008) "Aggregation of Attack Relations: A Social-

5 Cf. La introducción de Elster (1998).

Apéndice: Marcos argumentativos de Dung

Un *marco argumentativo* es un par $AF = \langle AR, \otimes \rangle$, donde AR es un conjunto de entidades llamadas *argumentos*, y \otimes es una relación binaria sobre AR , es decir, $\otimes \subseteq AR \times AR$, interpretada como derrota entre argumentos. La relación \otimes es arbitraria y no se supone, prima facie, que cumpla ninguna propiedad especial. El modelado de *extensiones* (i.e. conjuntos de argumentos ganadores en la interacción por derrotas) gira en torno a dos nociones básicas: la de “aceptabilidad” de un argumento y la de “admisibilidad” de un conjunto de argumentos. La idea es la de tomar aquellos argumentos que, si son derrotados, pueden ser defendidos mediante derrotas a sus derrotadores. Un argumento A es *aceptable* con respecto a un conjunto de argumentos S , si y sólo si:

$$\neg \exists B (B \otimes A \Rightarrow \exists C (C \in S \ \& \ C \otimes B)).$$

Un conjunto de argumentos S es *admisible* si y sólo si:

1. $\neg \exists A, B (A, B \in S \ \& \ A \otimes B)$ (S está libre de conflictos); y
2. todo argumento perteneciente a S es aceptable con respecto a S .

De las distintas nociones de extensión que Dung propone, mencionaremos la de *extensión preferida* (*preferred extension*), que busca capturar un comportamiento crédulo: una *extensión preferida* de AF es un conjunto máximamente (c.r. a la inclusión conjuntista) admisible de argumentos de AF . Por su parte, como semántica escéptica Dung propone la *extensión fundada* (*grounded extension*), que es el menor punto fijo del operador:

$$F(S) = \{A : A \text{ es aceptable con respecto a } S\}.$$