TEOREMA DA IMPOSSIBILIDADE DE ARROW

INTRODUÇÃO

Indivíduos como também grupos de indivíduos precisam tomar decisões em muitos diferentes contextos. Como indivíduos, temos de tomar decisões sobre como dividir nossa renda familiar entre diferentes objetivos. Uma firma tem de decidir entre várias diferentes coisas que ela necessita fazer a fim de competir efetivamente no mercado. Os governos necessitam tomar decisões sobre sua política externa, política doméstica, política fiscal e política monetária. Estudantes necessitam escolher os cursos que farão todo semestre. A lista de situações em que os indivíduos tem de tomar decisões de escolha é realmente bastante ampla.

Num problema típico de escolha individual, um indivíduo se depara com a situação de decidir o que escolher dentre um conjunto de escolhas alternativas disponíveis para ele. O seu ganho, recompensa ou "payoff" é uma conseqüência exclusiva da escolha que fizer. Diante das mesmas alternativas, dois indivíduos podem escolher bem diferentemente. Teria um indivíduo feito uma boa decisão e o outro uma má decisão? Obviamente, a resposta a essa pergunta reside no critério usado pelos indivíduos para avaliar decisões. Como é bem sabido, indivíduos têm diferentes objetivos e diversos interesses que podem afetar suas escolhas. Podemos predizer que um indivíduo racional tenha algum critério de escolha entre as alternativas, de acordo com o qual ele escolherá aquela que mais "prefere", isto é, uma alternativa que não lhe confere menor utilidade ou satisfação que qualquer outra escolha. Assim, por exemplo, poderia ser que ele preferisse trabalhar na firma b a trabalhar em qualquer outra firma porque o salário oferecido por b é maior do que o salário oferecido pelas outras firmas.

Num problema de escolha social o ganho de um indivíduo depende, em geral, não somente do que ele faz, mas também das decisões de outros indivíduos. Essas são situações tratadas na Teoria dos jogos. Existem vários indivíduos, cada um escolhendo uma ação do seu conjunto disponível de ações. As regras do jogo então determinam um payoff (uma recompensa, um ganho ou uma conseqüência) para cada indivíduo, que depende das ações escolhidas por todos os indivíduos. Portanto, dadas as regras do jogo, cada indivíduo racional quererá escolher a ação que lhe dará o maior ganho possível. Porém cada um tem de levar em conta que os outros estarão também tentando maximizar os seus ganhos. Cada indivíduo também tem crenças sobre as escolhas dos outros indivíduos. O "equilíbrio" ocorre quando todos os jogadores jogam racionalmente e as crenças de todos eles sobre o que os demais estão fazendo estão corretas. Às vezes este procedimento leva a uma competição entre os participantes, onde a alternativa escolhida por um indivíduo nem sempre corresponde àquela que ele prefere. Em outras situações isto leva a uma cooperação de benefício mútuo e, em geral, a uma combinação desses dois comportamentos extremos. **Em qualquer caso, as decisões tomadas por grupos de indivíduos diferem fundamentalmente das decisões tomadas por um único indivíduo.** É nesse sentido que a Teoria da Escolha Individual difere da Teoria dos Jogos. É justamente na natureza dessa diferença que se baseia a Teoria da Escolha Social que teve origem com o livro seminal de Keneth Arrow (1951, 1963). Por sua contribuição à Teoria Econômica, Arrow recebeu o prêmio Nobel de Economia em 1972, juntamente com Sir John Hicks.

2. RELAÇÃO DE PREFERÊNCIAS

Num problema de escolha individual existem um indivíduo j e um conjunto de alternativas X. O critério de escolha entre as alternativas é estabelecido por uma relação de preferência $\geq_j$ (ou R_j) sobre o conjunto X. Isto é, R_j é uma relação binária sobre X completa e transitiva. Assim, $x \gtrsim_j y$ (ou xR_jy) significa que o indivíduo j acha a opção x pelo menos tão boa quanto a opção y; lê-se "x é fracamente melhor que y para j".

A partir da relação de preferência $\geq$ definimos duas outras relações binárias, $>$ e $\approx$. Dizemos que $x >_j y$ (lê-se "j prefere x a y" ou "x é estritamente melhor que y para j") se $x \gtrsim_j y$ e não é verdade que $y \geq_j x$; dizemos que $x \approx_j y$ (lê-se "j é indiferente entre x e y" ou "x é pelo menos tão bom quanto y para j") se $x \gtrsim_j y$ e $y \gtrsim_j x$.

Por R_j ser completa, queremos dizer que dadas quaisquer alternativas a e b em X, o indivíduo j é sempre capaz de dizer se prefere a a b ou b a a ou se é indiferente entre ambas alternativas. Dizer que $\gtrsim_j$ é transitiva significa que se j prefere fracamente a a b e b a c então j prefere fracamente a a c, para quaisquer alternativas a, b e c em X. Vale observar que a transitividade de $\gtrsim_j$ implica na transitividade de $>_j$ e $\approx_j$.

É natural supor que o critério de escolha do indivíduo j seja consistente com a relação de preferências $\succsim_j$: se ele tem de escolher, por ex., entre x e y, e escolhe x então ele prefere x a y. Isto é dizer que o critério de esclha individual é *racional.*

Exigir que cada $\succsim_j$ seja completa é razoável. No entanto, **a transitividade é uma exigência forte** como sugere o exemplo a seguir.

Considere o conjunto de alternativas $X=\{x_0,x_1,x_2,\ldots,x_{100}\}$, onde x_0 é uma xícara de chá com uma colher de açúcar; x_1 é uma xícara de chá com menos 99/100 da colher de açúcar; …; x_{99} é uma xícara de chá com menos 1/100 da colher de açúcar; x_{100} é uma xícara de chá sem açúcar. É natural de se esperar que o indivíduo j seja indiferente entre x_0 e x_1; entre x_1 e x_2;…; entre x_{99} e x_{100}. Entretanto ele dificilmente será indiferente entre x_0 e x_{100}.

Por outro lado, o exemplo a seguir mostra que a **intransitividade, ou seja um critério de escolha baseado em preferências não transitivas, pode ser considerada irracional.** Considere o conjunto de alternativas $X=\{x,y,z\}$. Suponha que o indivíduo j prefere x a y, y a z, mas prefere z a x. Suponha que j tenha z. Como ele prefere y a z, estará disposto a pagar uma certa quantia, digamos, $0,10, para trocar z por y; novamente, como prefere x a y, estará disposto a pagar, digamos, $0,10, para trocar y por x. Finalmente, como prefere z a x, ele estará disposto a pagar uma certa quantia, seja $0,10, para trocar x por z. Assim, com uma preferência intransitiva, ele termina com a mesma alternativa que possuía, mas $0,30 mais pobre.

3. FUNÇÃO DO BEM ESTAR SOCIAL

Considere a seguinte situação de escolha. Um grupo de indivíduos (uma *sociedade*) deseja ir a um restaurante para comer uma pizza, que será repartida entre eles e deverá ser de um único sabor, para poupar dinheiro. Os tipos de pizza aceitáveis para qualquer um do grupo são: ***x*, *y* e *z***. Suponhamos que o restaurante escolhido altera com freqüência o seu cardápio de pizzas e ninguém do grupo tem informação sobre que tipos de pizza estarão disponíveis. Portanto, para efetuar suas escolhas, a sociedade deve definir suas preferências sobre ***x*, *y*** e ***z***. Assim, é pedido a cada indivíduo que entregue ao representante do grupo suas opções sobre os três tipos de pizza. Isto é, suas preferências sobre ***x*, *y*** e ***z*.**

A questão que surge naturalmente é:

Como efetuar as escolhas coletivas, racionalmente, levando em conta as escolhas individuais? Isto é, como definir uma relação de preferências para a sociedade, que expresse de forma racional as relações de preferências individuais?

Este é o assunto tratado no livro de Kenneth Arrow. Podemos interpretar que esta relação de preferências é uma pré-ordenação de todas as alternativas, de acordo com a qual a sociedade se compromete a agir. É desejável que um critério de escolha social seja consistente com coma relação binária e transitiva. Assim, um modelo de escolha social consiste de um conjunto de indivíduos (sociedade) **N={1,…,n};** um conjunto de alternativas **X**; para cada indivíduo j, um conjunto de preferências ρ_j e o conjunto das relações de preferências (completas e transitivas) da sociedade sobre X, ρ.

Uma **função do bem estar social (FBS)** agrega n pré-ordenações individuais (um perfil de preferências) de X numa única pré-ordenação (social) de X. Isto é,

$$\mathbf{F}: \rho_1 x \rho_2 x \ldots x \rho_n \quad \Rightarrow \quad \rho$$

$$\mathbf{R=(R_1,R_2,\ldots,R_n)} \qquad \mathbf{R_s}$$

Assim, **x F(R) y** significa que **o bem estar social não é mais baixo em x do que em y.**

O exemplo abaixo ilustra que uma das regras mais populares, a regra da maioria não é uma função do bem estar social.

Exemplo 1. (A regra da maioria não é uma FBS - Paradoxo de Condorcet - 1743 - 1794).

Considere o problema de escolha da pizza formulado anteriormente para uma sociedade formada pela união disjunta de três conjuntos de indivíduos *A, B* e *C,* com o mesmo número de elementos. As preferências dos agentes sobre o conjunto das pizzas X={x, y, z} são **estritas** e dadas por:

A	***B***	***C***
x	*y*	*z*
y	*z*	*x*
z	*x*	*y*

De acordo com a regra da maioria aplicada a este exemplo, a alternativa w é preferida à alternativa w' pela sociedade se o número de indivíduos que prefere w a w' é maior do que o número de pessoas que prefere o oposto. Esta regra não é uma FBS, pois aplicada ao perfil de preferências acima não produz uma relação binária transitiva. De fato, $x \geq_S y \geq_S z$ mas $z \geq_S x$. ■

É desejável que uma FBS induza critérios de escolha **racionais**. Como observado por Arrow, para se obter essa racionalidade, não basta consistência com uma relação de preferências. De fato, a regra da pluralidade é uma FBS. Porém não podemos considerar racional o critério de escolha induzido por esta regra. Vejamos:

Exemplo 2. (Regra da pluralidade dos votos). De acordo com esta regra, x é preferido a y pela sociedade se tem mais votos que y. Considere o problema de escolha da pizza formulado anteriormente. A sociedade é formada pela união disjunta de três conjuntos de indivíduos *A, B* e *C,* onde $|A|=3$, $|B|=|C|=2$. As preferências dos agentes sobre o conjunto de pizzas X={x, y, z} são dadas por:

A	***B***	***C***
x	*y*	*z*
y	*z*	*x*
z	*x*	*y*

Então $\boldsymbol{x \geq_s [y,z]}$**.** A pizza x é preferida a y e a z pela sociedade.

Considere agora a seguinte **cena no restaurante:** O garçom informa que o restaurante oferece as pizzas x, y e z. O representante do grupo pede a pizza x, a mais votada pelo grupo. O garçom vai até a cozinha e volta com a seguinte informação:

_ Queiram desculpar-nos, mas a pizza y não está boa hoje.

Os indivíduos de B, desencantados com y, mudam suas preferências para ≥':

B (≥')
z
x
y

Com as novas preferências de B a preferência do grupo muda para $\geq'_s$: $\boldsymbol{z} \boldsymbol{\geq'_s x} \boldsymbol{\geq'_s y}$.
Dessa forma, o representante do grupo informa ao garçom:

_ Pedimos a pizza x porque acreditávamos que y estava boa. Como y não está boa, desconsidere o pedido anterior e traga-nos a pizza z. ■

O princípio que a regra da pluralidade viola foi chamado por Arrow de:

A.1 - Independência das alternativas irrelevantes: Para todo par de alternativas x e y, as preferências entre x e y devem depender apenas de como as pessoas ordenam x em relação a y, e não de como ordenam as outras alternativas. Assim, se o conjunto de pessoas que preferem x a y num perfil de preferências R for igual ao conjunto de pessoas que preferem x a y num outro perfil de preferências R', então x F(R) y se e somente se x F(R') y.

Arrow listou mais três axiomas que considerou que uma "boa" FBS deve satisfazer:

A.2 - Unanimidade (ou princípio de Pareto): Para todo par de alternativas x e y, se $x \geq_j y$ para todo j=1,…,n, então $x \geq_s y$.

A.3- Domínio irrestrito: F é definida para qualquer conjunto $(\geq_1,\ldots,\geq_n)$ de preferências.

A.4- *F* deve ser não ditatorial: F é ditatorial se existe um indivíduo j (ditador) tal que para todo x e y em X, e para todo perfil de preferências $(\geq_1,\ldots,\geq_n) \in \rho_1 \mathrm{x} \ldots \mathrm{x} \rho_n$, se $x \geq_j y$ então $x\, F(\geq_1,\ldots,\geq_n)\, y$.

4. TEOREMA DA IMPOSSIBILIDADE DE ARROW

Nesta seção usaremos a seguinte notação: $x \geq_T y$ significa que $x \geq_j y$ para todo $j \in T$.

Definição 1: Dados x e y em X, e $T \subseteq N$, dizemos que ***T* determina *x* contra *y*** se para todo perfil de preferências em que $x \geq_T y$ e $y \geq_{N \setminus T} x$ então $x \geq_s y$, onde $\geq_s = F(\geq_T, \geq_{N \setminus T})$.

Observação: 1. Suponha que F satisfaz ao critério da unanimidade (A.2). Então, N determina x contra y, para todo x e y e ϕ não determina x contra y para nenhum par (x,y).

2. Suponha que F satisfaz A.1 e existe um perfil de preferências $(\geq,...,\geq)$ tal que:

T	$N \setminus T$
...	...
x	y
...	...
y	x

e $x \geq_s y$, onde $F(\geq,...,\geq) = \geq_s$. Então T determina x contra y.

Lema 1: Seja F uma FBS satisfazendo A-1,…,A-3. Suponha que $|X|>2$. Então existem x^* e y^* em X e $j^* \in N$ tal que j^* determina x^* contra y^*.

Demonstração: Seja $D = \{T \subseteq N;\ \exists\ x,y \in X$ tal que T dermina x contra $y\}$. Temos que $D \neq \phi$, pois $N \subseteq D$ por A.2 (unanimidade). Logo D tem um elemento minimal (conjunto parcialmente ordenado pela inclusão, finito e não vazio). Seja T^* tal elemento e x^* e y^* duas alternativas tais que T^* determina x^* contra y^*. Temos que $\phi \notin D$ pois senão existiriam x e y tal que se $y \geq_N x$ implicaria que $x \geq_s y$, contrariando A.2. Logo $|T^*| \geq 1$. Então existe $j^* \in N$ tal que $j^* \in T^*$. Vamos mostrar que $T^* = \{j^*\}$.

Suponhamos por absurdo que $|T^*| \geq 2$. Então $T^* = \{j^*\} \cup T^{**}$, onde $j^* \notin T^{**}$ e $|T^{**}| \geq 1$. Como $|X| \geq 3$ podemos tomar $z \neq x^*, y^*$. Por A-3 (domínio irrestrito) podemos considerar as seguintes preferências:

j^*	$T^* \setminus j^* = T^{**}$	$N \setminus T^*$
x^*	z	y^*
y^*	x^*	z

z	y^*	x^*

Temos que $x^* \geq_s y^*$ pois T^* determina x^* contra y^*. Se $z \geq_s y^*$ teríamos por A.1(independencia das alternativas irrelevantes) que $T^{**} \in D$, o que contraria a minimalidade de T^*. Como $\geq_s$ é completa implica que $y^* \geq_s z$. Portanto $x^* \geq_s y^* \geq_s z^*$ e por transitividade $x^* \geq_s z^*$. Por A.1 (independência das alternativas irrelevantes) segue que j^* determina x^* contra z, donde $\{j^*\} \in D$, contradição pois $\{j^*\} \subseteq T^*$, T^* é minimal e $|T^*| > 1$.■

Lema 2: Seja F uma FBS satisfazendo A-1,…,A-3. Suponha que $|X|>2$. Então $\{j^*\}$ é determinante.

Demonstração: Pelo Lema 1 existem x^* e y^* tal que j^* determina x^* contra y^*. Queremos mostrar que j^* determina w contra z para quaisquer z e w, com $z \neq w$. Seja $z \neq x^*, y^*$. ($\exists$ z, pois $|X| \geq 3$). Por A-3 (domínio irrestrito) podemos considerar as preferências:

j^*	$N \backslash j^*$
x^*	y^*
y^*	z
z	x^*

$x^* \geq_s y^*$ pois j^* determina x^* contra y^*; $y^* \geq_s z$, por A.2; $x^* \geq_s z$, por transitividade. Então, por A.1 (independência das alternativas irrelevantes),

j^* determina x^* contra z. (1)

Tome, agora, $w \neq z, x^*$. Considere as preferências:

j^*	$N \backslash j^*$
w	z
x^*	w
z	x^*

$w \geq_s x^*$ por A.2 (unanimidade); $x^* \geq_s z$, pois j determina x^* contra z por (1); $w \geq_s z$ por transitividade. Então, por A.1 (independência das alternativas irrelevantes), j determina w contra z. Como w e z são arbitrários o lema está demonstrado. ■

Lema 3: Seja F uma FBS satisfazendo A-1,…,A-3. Suponha que $|X|>2$. Se j é determinante então j é um ditador.

Demonstração: Se j é determinante então j faz prevalecer sua opinião sempre que os outros estão contra. Temos que mostrar que j sempre faz prevalecer sua opinião. Então, dados $x, y, z \in X$, considere:

j	$N \backslash j$	
x	z	z
z	x	y
y		

Por A-3 (domínio irrestrito) estas preferências estão no domínio de F. Então $x \geq_s z$, pois j é determinante; $z \geq_s y$, por A.2 (unanimidade); $x \geq_s y$, por transitividade; $x \geq'_s y$ para toda $(\geq'_{1,\ldots,} \geq'_n)$ que mantém as preferências entre x e y. Logo j é um ditador. ■

Arrow provou que se o número de alternativas for maior do que dois, então é impossível definir uma FBS que induza um critério de escolha racional.

Teorema 1 (Teorema da impossibilidade de Arrow): Se $|X|>2$ então não existe nenhuma FBS que satisfaz A-1,…,A-4.

Demonstração: Assumamos que F é uma FBS satisfazendo A-1,…,A-4. Pelos Lemas 1 e 2 existe j^* determinante. Pelo Lema 3, j^* é ditador e portanto a função F não satisfaz A-4, contradição. ■

5. OBSERVAÇÕES FINAIS

O cerne da Teoria da Escolha Social trata de problemas de escolha, melhor descritos como uma situação de votação: um único candidato é escolhido de um conjunto de candidatos elegíveis, tendo-se em conta como dados somente as preferências ordinais dos votantes. No final da década de 60 cresceu o interesse pelas propriedades estratégicas das regras de votação, e a existência ou não de uma regra de votação *à prova de estratégia* (notadamente uma regra em que votar no candidato preferido é a melhor resposta para todo votante, independentemente do que os outros jogadores votem) que não desse inteiro poder de decisão a um votante ditador, passou, daí em diante, a ser alvo de investigação entre os autores. Essa questão foi respondida negativamente por Gibbard (1973) e Satterthwaite (1975) em dois artigos quase simultâneos. Este teorema teve profunda influência sobre a então emergente literatura sobre compatibilidade de incentivos e desenho de mecanismos. A teoria estratégica de votação também inspirou a teoria de implementação.

REFERÊNCIAS

Arrow, Kenneth J., Social Choice and Individual Values. New Haven, Yale University Press, 1951, 1963.

Este é o trabalho original: um dos raros exemplos de uma peça de um trabalho realmente seminal, que é entretanto de fácil leitura.

Sen, Amartya K., Collective Choice and Social Welfare, San Francisco, Holden-Day, 1970.

Este livro oferece um bom tratamento da teoria de escolha social até 1970.

Kelley, Jerry S., Arrow Impossibility Theorems, New York, Academic Press, 1978.

Este livro inclui alguns importantes desenvolvimentos não incluídos no livro de Sen. Porém o tratamento não é de leitura fácil.