IZBORNI PARADOKSI

MILA BOTIĆ I VEDRAN KRČADINAC

Sažetak. Objašnjavamo matematički model za izborni proces. Iskazujemo teoreme nemogućnosti, prema kojima određeni "demokratski" zahtjevi na metodu određivanja pobjednika izbora dovode do postojanja "diktatora".

1. Uvod

Teorija društvenog izbora (eng. *social choice theory*) je disciplina društvenih znanosti na granici između ekonomije, politologije i filozofije. Bavi se načinom donošenja kolektivnih odluka na temelju preferencija pojedinaca. Početkom moderne teorije društvenog izbora smatra se knjiga *Social choice and individual values* [1] američkog ekonomista Kennetha J. Arrowa, dobitnika Nobelove nagrade za ekonomiju 1972. godine. Među ranijim autorima koji su pisali o sličnoj tematici su matematičari Jean-Charles de Borda (1733.-1799.), Nicolas Caritat de Condorcet (1743.-1794.), Pierre-Simon Laplace (1749.-1827.) i Charles Lutwidge Dodgson (1832.-1898.), poznatiji kao Lewis Carroll – autor "Alise u zemlji čudesa". Teorija se i danas razvija, o čemu svjedoče znanstveni časopisi *Social Choice and Welfare* (ISSN 0176-1714) i *Public Choice* (ISSN 0048-5829) iz srodne discipline, teorije javnog izbora.

U modernim znanstvenim radovima iz ekonomije u velikoj se mjeri koristi precizan matematički jezik. Cilj ovog članka je prezentirati neke matematičke aspekte teorije društvenog izbora. Objašnjavamo osnovne definicije i ideje koje stoje iza njih te iskaze takozvanih teorema nemogućnosti, koji se mogu opisati kao izborni paradoksi. Riječ je o zahtjevima na metodu određivanja pobjednika izbora koji djeluju pravedno i razumno te ćemo ih stoga zvati "aksiomima demokracije". Međutim, iz zahtjeva slijedi postojanje "diktatora", tj. birača čiji glas jednoznačno određuje ishod izbora neovisno o glasovima ostalih birača. Dokaze teorema u ovom članku preskačemo i upućujemo zainteresirane čitatelje na literaturu u kojoj se nalaze. Na kraju svake cjeline navodimo zadatke

Date: 12.9.2011.

koji služe boljem razumijevanju izloženih pojmova. Rješenja zadataka dajemo na kraju članka.

2. Kako rangirati kandidate?

Na izborima biramo između konačnog skupa kandidata, stranaka ili opcija. Na predsjedničkim izborima biramo jednog kandidata, na parlamentarnim izborima biramo listu kandidata, a na referendumu odlučujemo između nekoliko opcija. Sva tri slučaja tretirat ćemo ravnopravno. Elemente skupa K koji su predmet izbora zvat ćemo *kandidatima*, a broj elemenata u K označavat ćemo s k.

Na većini izbora u Republici Hrvatskoj birači se trebaju izjasniti za točno jednog kandidata. Teorija društvenog izbora proučava općenitije izborne sustave, u kojima se uzima u obzir redoslijed u kojem birači rangiraju kandidate, a ne samo njihov prvi izbor. Jedan takav izborni sustav je takozvani *instant runoff voting* ili *alternative vote* [11]. Koristi se na nekim izborima u Australiji, Kanadi, SAD-u, Velikoj Britaniji i drugdje. Prvo ćemo objasniti što se točno podrazumijeva pod rangiranjem kandidata.

Svaki od birača uspostavlja uređaj na skupu kandidata koji odgovara njegovim preferencijama. Bitan je samo redoslijed kandidata u tom uređaju, a ne stupanj potpore pojedinim kandidatima. Naprimjer, ako se uređaj uspostavlja dodjeljivanjem bodova, birač uspostavlja identične uređaje time što kandidatima x, y, z redom dodijeli bodove 10, 20, 30 ili 5, 10, 100.

Formalno, uređaj ρ je binarna relacija na skupu kandidata K, tj. podskup Kartezijeva produkta $\rho \subseteq K \times K$. Za parove kandidata koji su u relaciji pišemo $x \,\rho\, y$ umjesto $(x, y) \in \rho$. U sljedećoj definiciji navedeni su zahtjevi koji se postavljaju na razne vrste relacija uređaja.

Definicija 2.1. *Za relaciju $\rho \subseteq K \times K$ kažemo da je:*

- refleksivna*, ako za svaki $x \in K$ vrijedi $x \,\rho\, x$;*
- irefleksivna*, ako za svaki $x \in K$ ne vrijedi $x \,\rho\, x$ (što zapisujemo kao $x \not\rho\, x$);*
- antisimetrična*, ako iz $x \,\rho\, y$ i $y \,\rho\, x$ slijedi $x = y$;*
- tranzitivna*, ako iz $x \,\rho\, y$ i $y \,\rho\, z$ slijedi $x \,\rho\, z$;*
- potpuna*, ako za sve $x, y \in K$ vrijedi $x \,\rho\, y$ ili $y \,\rho\, x$ ili $x = y$.*

Relacija "manje ili jednako" $\leq$ na skupu realnih brojeva $\mathbb{R}$ ili nekom njegovom podskupu je refleksivna, antisimetrična, tranzitivna i potpuna. Relacija "strogo manje" $<$ ima ista svojstva, osim što je irefleksivna umjesto refleksivna. Takve relacije nazivamo relacijama *totalnog uređaja* ili *linearnog uređaja*.

Općenitije su relacije *parcijalnog uređaja*, koje ispunjavaju zahtjeve refleksivnosti, antisimetričnosti i tranzitivnosti, ali ne moraju biti potpune. Primjeri su djeljivost na skupu prirodnih brojeva $\mathbb{N}$ i inkluzija $\subseteq$ (relacija "biti podskup") na nekoj familiji skupova. I ovdje refleksivnost možemo zamijeniti s irefleksivnosti, prijelazom na strogu inkluziju skupova $\subset$.

U parcijalno uređenom skupu mogu postojati neusporedivi elementi, tj. različiti elementi $x, y \in K$ takvi da je $x \not\rho y$ i $y \not\rho x$. Naprimjer, prirodni brojevi 2 i 3 su neusporedivi obzirom na relaciju djeljivosti, a skupovi $\{1, 2\}$ i $\{2, 3\}$ su neusporedivi obzirom na inkluziju. Neusporedivost kandidata na izborima znači da su jednako rangirani. Zato parcijalni uređaj na K treba imati dodatno svojstvo: neusporedivost mora biti tranzitivna. Takve relacije nazivamo *strogim slabim uređajima.*

Definicija 2.2. *Za relaciju $\rho \subseteq K \times K$ kažemo da je* strogi slabi uređaj *na K ako je irefleksivna, antisimetrična, tranzitivna i ako je neusporedivost obzirom na tu relaciju tranzitivna.*

U teoriji društvenog izbora preferencije birača zadaju se relacijama strogog slabog uređaja. U njima možemo identificirati međusobno neusporedive kandidate kao ekvivalentne i dobiti totalni uređaj na klasama ekvivalencije. Ponekad se radi jednostavnosti ne dozvoljava jednako rangiranje kandidata, tj. zahtijeva se totalni uređaj na K. Tako postupamo i u ovom članku: pretpostavljamo da svaki birač zadaje irefleksivnu, antisimetričnu, tranzitivnu i potpunu relaciju na K. Skup svih takvih relacija označavamo s $\mathcal{T}$, a pojedine relacije iz $\mathcal{T}$ označavamo s $\prec$ (da bismo ih razlikovali od standardnog uređaja $<$ na skupu realnih brojeva $\mathbb{R}$).

Ako vrijedi $x \prec y$, smatramo da je kandidat x bolje rangiran od kandidata y. Naravno, moguć je i obrnuti dogovor, baš kao što relaciju "manje" $<$ na $\mathbb{R}$ možemo zamijeniti s relacijom "veće" $>$. Te relacije imaju ista svojstva: obje su irefleksivne, antisimetrične, tranzitivne i potpune.

Totalne uređaje možemo identificirati s permutacijama skupa K, tj. bijekcijama sa $\{1, \ldots, k\}$ na K. Iz toga slijedi da totalnih uređaja na K ima točno $k!$. Inverzna funkcija neke permutacije je bijekcija $u : K \to \{1, \ldots, k\}$ koju tumačimo kao funkciju korisnosti (eng. *utility function*): vrijednost $u(x)$ je mjesto na kojem je rangiran kandidat x. Općenitije, bilo koja funkcija $u : K \to \mathbb{R}$ definira jedan strogi slabi uređaj na K sa $x \prec y \iff u(x) < u(y)$. Totalne uređaje dobivamo od injektivnih funkcija $u : K \to \mathbb{R}$.

Zadaci.

2.1. Dokažite da iz irefleksivnosti i tranzitivnosti binarne relacije slijedi njezina antisimetričnost. Prema tome, uvjet antisimetričnosti možemo izostaviti iz definicije 2.2.

2.2. Nađite primjer relacije parcijalnog uređaja u kojoj neusporedivost nije tranzitivna.

2.3. Neka je $\prec$ tranzitivna relacija. Dokažite da je tada tranzitivnost neusporedivosti ekvivalentna sljedećem zahtjevu: ako vrijedi $x \prec y$, onda za svaki z vrijedi $x \prec z$ ili $z \prec y$.

2.4. Dokažite da za svaki strogi slabi uređaj $\prec$ na skupu K postoji funkcija $u : K \to \mathbb{R}$ takva da vrijedi $x \prec y \iff u(x) < u(y)$.

2.5. Koliko ima relacija strogog slabog uređaja na k-članom skupu?

3. Kako odrediti pobjednika?

Pretpostavimo da na izbore izlazi n birača; identificiramo ih s prirodnim brojevima $1, 2, \ldots, n$. Svaki birač zadaje relaciju totalnog uređaja na skupu kandidata K kojom izražava svoje preferencije. Tako dobivamo uređenu n-torku $(\prec_1, \ldots, \prec_n) \in \mathcal{T}^n$ koju nazivamo *profilom* ili *n-torkom individualnih preferencija.* Glavni zadatak teorije društvenog izbora je na temelju profila odrediti pobjednika izbora, odnosno *društvenu preferenciju*: totalni uređaj $\prec$ koji predstavlja preferencije društva kao cjeline. Nicolas de Condorcet je 1785. godine demonstrirao jedan izborni paradoks i pokazao da zadatak nije jednostavan.

Pretpostavimo da na izborima glasa samo tri birača 1, 2, 3 i da se izjašnjavaju o tri kandidata x, y, z. Neka su preferencije prvog birača $x \prec_1 y \prec_1 z$, drugog birača $y \prec_2 z \prec_2 x$, a trećeg birača $z \prec_3 x \prec_3 y$. Kandidat x ne bi trebao pobijediti na izborima jer većina birača preferira kandidata z: vrijedi $z \prec_2 x$ i $z \prec_3 x$. Kandidat z također ne bi trebao pobijediti jer prvi i drugi birač preferiraju kandidata y. Niti y ne bi trebao pobijediti, zbog $x \prec_1 y$ i $x \prec_3 y$. Ovaj profil individualnih preferencija je paradoksalan jer ne dopušta takozvanog *Condorcetovog pobjednika.* To je kandidat koji bi u drugom krugu glasanja pobijedio svakog od preostalih kandidata, ako bi birači glasali konzistentno s preferencijama izraženim u prvom krugu.

Definicija 3.1. *Neka je $(\prec_1, \ldots, \prec_n) \in \mathcal{T}^n$ profil individualnih preferencija na K. Za kandidata $x \in K$ kažemo da je* Condorcetov pobjednik *zadanog profila, ako za svaki $y \in K$, $y \neq x$ vrijedi da je broj birača* *koji preferiraju x veći od broja birača koji preferiraju y:*

$$\mathrm{card}\{i \mid x \prec_i y\} > \mathrm{card}\{j \mid y \prec_j x\}.$$

Condorcetov paradoks pokazuje da za neke profile ne postoji Condorcetov pobjednik.

U ambicioznijoj verziji izbornog problema treba odrediti društvenu preferenciju za sve parove kandidata, a ne samo pobjednika. Godine 1770. Jean-Charles de Borda predložio je metodu za određivanje uređaja $\prec$ na temelju profila $(\prec_1, \ldots, \prec_n)$. Kandidati dobivaju bodove ovisno o mjestu na kojem su rangirani u individualnim preferencijama $\prec_i$. Bodovi se sumiraju i na osnovu toga se uspostavlja društvena preferencija $\prec$. Postoji više varijanti Bordine metode. Borda ju je predložio za izbor članova francuske Akademije znanosti i tamo se koristila od 1784. do 1800. [10]. Danas se sličan sistem koristi za izbor zastupnika nacionalnih manjina u parlamentu Slovenije te za izbor pjesme Eurovizije. Obično bolje rangirani kandidati dobivaju više bodova, no mi ćemo pretpostaviti suprotno.

Točnije, neka su $u_i : K \to \{1, \ldots, k\}$ funkcije korisnosti koje odgovaraju individualnim preferencijama $\prec_i$, za $i = 1, \ldots, n$. Definiramo funkciju $u : K \to \mathbb{R}$, $u(x) = \sum_{i=1}^{n} u_i(x)$. Društvena preferencija $\prec$ je strogi slabi uređaj dobiven od te funkcije, tj. definiran s $x \prec y \iff u(x) < u(y)$. Nije isključeno da se vrijednosti funkcije u podudaraju za različite kandidate, a tada su ti kandidati neusporedivi obzirom na relaciju $\prec$. Ako želimo totalni uređaj na K, možemo se dogovoriti da se u tom slučaju gleda preferencija unaprijed zadanog birača[1] ili da se izjednačeni kandidati rangiraju na neki drugi način.

Općenito, teorija društvenog izbora proučava funkcije $f : \mathcal{T}^n \to K$ i $F : \mathcal{T}^n \to \mathcal{T}$. Funkcije s domenom i kodomenom kao f zovemo *funkcijama društvenog izbora*. One svakom profilu $(\prec_1, \ldots, \prec_n)$ pridružuju pobjednika izbora $f(\prec_1, \ldots, \prec_n) \in K$. Arrow [1] je proučavao funkcije s domenom i kodomenom kao F, koje profilu $(\prec_1, \ldots, \prec_n)$ pridružuju društvenu preferenciju $F(\prec_1, \ldots, \prec_n) \in \mathcal{T}$. Zvao ih je *funkcijama društvenog blagostanja* (eng. *social welfare function*), a koristi se i naziv *ustav* (*constitution*).

Bordina metoda definira funkciju društvenog blagostanja F_B, odnosno funkciju društvenog izbora f_B (za pobjednika uzimamo kandidata x za kojeg je $u(x)$ minimalno). Postoje mnogi drugi primjeri takvih funkcija. Najjednostavniji primjer je projekcija na koordinatu $d \in \{1, \ldots, n\}$:

$$F_d(\prec_1, \ldots, \prec_n) = \prec_d,$$

[1] U Pravilniku PMF–Matematičkog odsjeka stoji odredba "u slučaju izjednačenog broja glasova odlučuje glas pročelnika".

odnosno $f_d(\prec_1, \ldots, \prec_n) = x$ ako je x najbolje rangirani kandidat u $\prec_d$. To znači da preferencije birača d potpuno određuju rezultat izbora, a preferencije ostalih birača uopće ne utječu na rezultat. Takvog birača nazivamo *diktatorom.*

Diktatorske funkcije F_d i f_d su krajnje nedemokratične i cijeli izborni proces dovode do apsurda. Zato se u teoriji društvenog izbora postavljaju zahtjevi koje trebaju zadovoljavati funkcije društvenog blagostanja i funkcije društvenog izbora. Takve zahtjeve zvat ćemo *aksiomima demokracije.*

Zadaci.

3.1. Dokažite da je Condorcetov pobjednik jedinstven, ako postoji.

3.2. Neka je zadana funkcija društvenog blagostanja $F : \mathcal{T}^n \to \mathcal{T}$. Tada na prirodan način možemo definirati funkciju društvenog izbora $f : \mathcal{T}^n \to K$, tako da $f(\prec_1, \ldots, \prec_n)$ bude najbolje rangirani kandidat u uređaju $F(\prec_1, \ldots, \prec_n)$. Možemo li na taj način dobiti svaku funkciju društvenog izbora $f : \mathcal{T}^n \to K$?

4. Aksiomi demokracije

Prvi od zahtjeva na funkciju društvenog izbora f jest da svaki od kandidata može pobijediti, tj. da je f surjekcija.

Aksiom $(\mathbf{a_1})$**.** *Za svaki* $x \in K$ *postoji profil* $(\prec_1, \ldots, \prec_n) \in \mathcal{T}^n$ *takav da je* $f(\prec_1, \ldots, \prec_n) = x$.

Nešto jači zahtjev je takozvani *Paretov uvjet.*

Aksiom $(\mathbf{a_1'})$**.** *Ako je* $x \in K$ *najbolje rangirani kandidat u svakoj od individualnih preferencija* $\prec_1, \ldots, \prec_n$*, onda je* $f(\prec_1, \ldots, \prec_n) = x$.

Vilfredo Pareto (1848.-1923.) bio je talijanski ekonomist, sociolog i filozof. Zaslužan je za širenje matematičke notacije i matematičkih tehnika u radovima iz ekonomije. Očito iz Paretova uvjeta slijedi surjektivnost, a uz neke dodatne zahtjeve aksiomi (a_1) i (a_1') su ekvivalentni (vidi zadatak 4.1). Surjektivnost možemo zahtijevati i od funkcije društvenog blagostanja F.

Aksiom $(\mathbf{A_1})$**.** *Za svaki totalni uređaj* $\prec$ *na* K *postoji profil* $(\prec_1, \ldots, \prec_n) \in \mathcal{T}^n$ *takav da je* $F(\prec_1, \ldots, \prec_n) = \prec$.

Arrow u knjizi [1] postavlja nešto blaži zahtjev: u društvenoj preferenciji svaki par kandidata može biti rangiran u bilo kojem redoslijedu.

Aksiom $(\mathbf{A_1'})$**.** *Za svaka dva kandidata* $x, y \in K$ *postoji profil* $(\prec_1, \ldots, \prec_n) \in \mathcal{T}^n$ *takav da za društvenu preferenciju* $\prec = F(\prec_1, \ldots, \prec_n)$ *vrijedi* $x \prec y$.

Arrow ovaj zahtjev zove *uvjetom suvereniteta građana.* Možemo ga malo ojačati u skladu s Paretovim uvjetom:

Aksiom ($\mathbf{A_1''}$). *Ako u svakoj od individualnih preferencija vrijedi* $x \prec_i y$, $i = 1, \ldots, n$, *onda u društvenoj preferenciji* $\prec = F(\prec_1, \ldots, \prec_n)$ *također vrijedi* $x \prec y$.

Druga vrsta zahtjeva na funkcije društvenog izbora i blagostanja je *monotonost.* Pojednostavljeno, monotonost znači da se položaj kandidata ne bi trebao pogoršati ako ga birači rangiraju bolje nego ranije. Neka je $x \in K$ kandidat i $\prec, \prec' \in \mathcal{T}$ dva totalna uređaja na K. Kažemo da je x bolje rangiran u $\prec'$ nego u $\prec$ ako vrijedi:

(1) ako za kandidata y vrijedi $x \prec y$, onda vrijedi $x \prec' y$;
(2) za sve kandidate y, z različite od x vrijedi $y \prec z$ ako i samo ako vrijedi $y \prec' z$.

To znači da je x u $\prec'$ na istom ili boljem položaju nego u $\prec$, a redoslijed ostalih kandidata je isti. Kažemo da je x bolje rangiran u profilu $P' = (\prec_1', \ldots, \prec_n')$ nego u profilu $P = (\prec_1, \ldots, \prec_n)$ ako to vrijedi za svaki par individualnih preferencija $\prec_i'$, $\prec_i$, $i = 1, \ldots, n$. Tu činjenicu zapisujemo ovako: $P' <_x P$. Arrow [1] sljedeći zahtjev zove *pozitivna asociranost* društvene i individualnih preferencija. Mi ćemo ga zvati *monotonost* funkcije društvenog blagostanja.

Aksiom ($\mathbf{A_2}$). *Neka su* $x, y \in K$ *kandidati,* $P, P' \in \mathcal{T}^n$ *profili i* $\prec = F(P)$, $\prec' = F(P')$ *odgovarajuće društvene preferencije. Ako vrijedi* $x \prec y$ *i* $P' <_x P$, *onda vrijedi* $x \prec' y$.

Oslabljivanjem pretpostavke dobivamo jači aksiom. Za profile $P = (\prec_1, \ldots, \prec_n)$ i $P' = (\prec_1', \ldots, \prec_n')$ pišemo $P' \leq_x P$ ako za svaki $i \in \{1, \ldots, n\}$ vrijedi

(1) ako za kandidata y vrijedi $x \prec_i y$, onda vrijedi $x \prec_i' y$.

Dakle, x je u svakoj komponenti od P' na istom ili boljem položaju nego u odgovarajućoj komponenti od P, ali redoslijed ostalih kandidata ne mora biti isti. Sljedeći aksiom zovemo *jaka monotonost* funkcije društvenog blagostanja ili *jaka pozitivna asociranost* društvene i individualnih preferencija.

Aksiom ($\mathbf{A_2'}$). *Neka su* $x, y \in K$ *kandidati,* $P, P' \in \mathcal{T}^n$ *profili i* $\prec = F(P)$, $\prec' = F(P')$ *odgovarajuće društvene preferencije. Ako vrijedi* $x \prec y$ *i* $P' \leq_x P$, *onda vrijedi* $x \prec' y$.

Aksiome monotonosti i jake monotonosti možemo izreći i za funkcije društvenog izbora.

Aksiom $(\mathbf{a_2})$. *Ako za kandidata $x \in K$ i profile $P, P' \in \mathcal{T}^n$ vrijedi $x = f(P)$ i $P' <_x P$, onda vrijedi $x = f(P')$.*

Aksiom $(\mathbf{a_2'})$. *Ako za kandidata $x \in K$ i profile $P, P' \in \mathcal{T}^n$ vrijedi $x = f(P)$ i $P' \leq_x P$, onda vrijedi $x = f(P')$.*

Arrow [1] postavlja još jedan zahtjev na funkciju društvenog blagostanja, takozvanu *nezavisnost od nevažnih alternativa* (eng. *independence of irrelevant alternatives*).

Aksiom $(\mathbf{A_3})$. *Neka su $P = (\prec_1, \ldots, \prec_n)$, $P' = (\prec_1', \ldots, \prec_n') \in \mathcal{T}^n$ profili i $S \subseteq K$ podskup kandidata. Ako se restrikcije individualnih preferencija $\prec_i|_S = \prec_i'|_S$ podudaraju za svaki $i = 1, \ldots, n$, onda se restrikcije društvenih preferencija $\prec = F(P)$, $\prec' = F(P')$ također podudaraju: $\prec|_S = \prec'|_S$.*

Restrikcija relacije $\prec \subseteq K \times K$ na podskup $S \subseteq K$ je relacija $\prec|_S = \prec \cap (S \times S)$. Ako je $\prec$ totalni uređaj na K, onda je restrikcija $\prec|_S$ totalni uređaj na S. Analogna tvrdnja vrijedi za stroge slabe uređaje. Smisao aksioma (A_3) jest da odustajanje nekih kandidata ne bi trebalo utjecati na izborni položaj preostalih kandidata. U suprotnom neki kandidati mogu manipulirati izborima na način da potiču kandidiranje "lažnih kandidata", kojima je cilj "pokupiti" glasove njihovih takmaca i odustati. Sličan aksiom mogao bi se formulirati za funkciju društevnog izbora f, ali se u tom kontekstu obično promatra druga vrsta manipulacije, koju provode birači.

Sjetimo se da birači rangiraju sve kandidate. Funkcija društvenog izbora $f : \mathcal{T}^n \to K$ određuje samo jednog pobjednika, a svi ostali kandidati su gubitnici izbora. U tom kontekstu birači često reagiraju tako da svojeg favorita rangiraju najbolje, a kandidate koje doživljavaju kao njegove konkurente rangiraju lošije od svojih stvarnih preferencija. Takvo ponašanje naziva se *strateškim glasanjem* i smatra se nedostatkom izbornog sustava koji ga potiče. Idući aksiom naziva se *otpornost na strateško glasanje.*

Aksiom $(\mathbf{a_4})$. *Neka je $P = (\prec_1, \ldots, \prec_n) \in \mathcal{T}^n$ profil, $i \in \{1, \ldots, n\}$ birač te $P' = (\prec_1, \ldots, \prec_i', \ldots, \prec_n)$ profil kojeg dobijemo zamjenom i-te koordinate od P s uređajem $\prec_i' \in \mathcal{T}$. Ako je $f(P) \neq f(P')$, onda je $f(P) \prec_i f(P')$.*

Uređaj $\prec_i$ tumačimo kao stvarnu preferenciju i-tog birača. Negacija aksioma (a_4) glasi: postoji profil P i birač i takav da mu se isplati zamijeniti svoju stvarnu preferenciju s uređajem $\prec_i'$. Tada će pobjednik izbora $f(P')$ biti bolje rangiran od $f(P)$ u stvarnoj preferenciji $\prec_i$. To znači da birač i ima razloga za strateško glasanje – u aksiomu zahtijevamo da to ne bude istina niti za jednog birača.

Zadaci.

4.1. Dokažite da iz surjektivnosti (a_1) i jake monotonosti (a'_2) funkcije društvenog izbora slijedi Paretov uvjet (a'_1).

4.2. Koje implikacije vrijede, a koje ne vrijede između aksioma (A_1), (A'_1) i (A''_1)? Koje implikacije vrijede za jako monotone funkcije društvenog blagostanja, tj. pod dodatnom pretpostavkom (A'_2)?

5. Teoremi nemogućnosti

Promotrimo koje od aksioma zadovoljava Bordina funkcija društvenog blagostanja F_B. Funkcija je očito surjektivna: za bilo koji totalni uređaj $\prec \in \mathcal{T}$, ako se sve individualne preferencije podudaraju s $\prec$, onda to vrijedi i za društvenu preferenciju $F_B(\prec, \ldots, \prec) = \prec$. Prema tome, ispunjen je aksiom (A_1). Ispunjen je i aksiom (A''_1), iz kojeg direktno slijedi (A'_1): ako za individualne preferencije vrijedi $x \prec_i y$, onda za odgovarajuće funkcije korisnosti vrijedi $u_i(x) < u_i(y)$, $i = 1, \ldots, n$. Društvenu preferenciju određuje suma $u(x) = \sum_{i=1}^n u_i(x) < \sum_{i=1}^n u_i(y) = u(y)$, pa je $x \prec y$ i za $\prec = F_B(\prec_1, \ldots, \prec_n)$.

Bordina funkcija društvenog blagostanja F_B je monotona, tj. zadovoljava aksiom (A_2). Iz $x \prec y$ slijedi $\sum_{i=1}^n u_i(x) < \sum_{i=1}^n u_i(y)$. Ako za profile P, P' vrijedi $P' <_x P$, onda je $u'_i(x) \leq u_i(x)$, za $i = 1, \ldots, n$. Nadalje, $u'_i(y)$ ostao je isti kao $u_i(y)$ ili se povećao za jedan ako je x "prestigao" y u i-toj individualnoj preferenciji $\prec'_i$. U svakom slučaju vrijedi $u'_i(y) \geq u_i(y)$, za $i = 1, \ldots, n$. Zato imamo $\sum_{i=1}^n u'_i(x) \leq \sum_{i=1}^n u_i(x) < \sum_{i=1}^n u_i(y) \leq \sum_{i=1}^n u'_i(y)$, iz čega slijedi $x \prec' y$.

Funkcija F_B ipak ne zadovoljava jaku monotonost (A'_2) i nezavisnost od nevažnih alternativa (A_3). Kao protuprimjer za (A_3) promotrimo slučaj s tri birača i četiri kandidata x, y, z, w. Kandidati su u profilu $(\prec_1, \prec_2, \prec_3)$ rangirani na sljedeći način:

	x	y	z	w
u_1	1	2	3	4
u_2	1	2	3	4
u_3	2	1	3	4
u	4	5	9	12

Individualne preferencije zadane su funkcijama korisnosti u_1, u_2, u_3, a u zadnjem retku tablice je suma $u = u_1 + u_2 + u_3$ koja određuje društvenu preferenciju. Drugi profil $(\prec'_1, \prec'_2, \prec'_3)$ zadan je sljedećom tablicom:

	x	y	z	w
u'_1	1	2	3	4
u'_2	1	2	3	4
u'_3	4	1	2	3
u'	6	5	8	11

Restrikcije individualnih preferencija na podskup kandidata $\{x, y\}$ podudaraju se u oba profila. Za prva dva birača je $x \prec_1 y$ i $x \prec_2 y$, a za trećeg birača je $y \prec_3 x$. Isti odnosi vrijede u uređajima $\prec'_1$, $\prec'_2$ i $\prec'_3$. Međutim, restrikcije društvenih preferencija $\prec = F_B(\prec_1, \prec_2, \prec_3)$ i $\prec' = F_B(\prec'_1, \prec'_2, \prec'_3)$ se ne podudaraju: vrijedi $x \prec y$ i $y \prec' x$.

Ovaj primjer odgovara situaciji u kojoj većina birača preferira kandidata x nad y. Međutim, manjina koja preferira y izrazito je nesklona kandidatu x. Ako u izborima sudjeluju samo x i y, pobjeđuje x. Ako pak sudjeluju i druga dva kandidata z i w, pobjeđuje y jer njegovi birači stavljaju kandidata x na zadnje mjesto.

Bordina funkcija društvenog izbora f_B očito zadovoljava Paretov uvjet (a'_1), iz čega odmah slijedi surjektivnost (a_1). Slično kao za F_B pokazuje se da je f_B monotona (vrijedi (a_2)), ali nije jako monotona (ne vrijedi (a'_2)). Osim toga f_B nije otporna na strateško glasanje, tj. ne zadovoljava aksiom (a_4). To se vidi iz istog primjera kao malo prije: trećem biraču isplati se umjesto stvarne preferencije $\prec_3$ na glasanju iskazati $\prec'_3$ jer tada pobjeđuje njegov favorit y.

Za razliku od Bordinih funkcija, diktatorske funkcije društvenog blagostanja F_d i društvenog izbora f_d zadovoljavaju sve aksiome demokracije. Teoremi nemogućnosti tvrde da je situacija još paradoksalnija: ako postoje bar tri kandidata, diktatorske funkcije su jedine koje zadovoljavaju sve aksiome!

Teorem 5.1 (Arrow). *Neka je $F : \mathcal{T}^n \to \mathcal{T}$ funkcija društvenog blagostanja koja zadovoljava aksiome (A'_1), (A_2) i (A_3). Ako je $k \geq 3$, onda postoji $d \in \{1, \ldots, n\}$ takav da je $F = F_d$.*

Ovo je najvažniji rezultat iz [1]. Arrow je uz spomenute aksiome imao aksiom da funkcija društvenog blagostanja ne smije biti diktatorska. U tom kontekstu teorem tvrdi da nije moguća funkcija društvenog blagostanja koja zadovoljava sve aksiome – odatle naziv "teorem nemogućnosti"[2].

U literaturi postoje mnogi alternativni dokazi i modifikacije Arrowljeva teorema. Jedan kratak dokaz dan je u knjizi [3]. Među najpoznatijim srodnim teoremima je takozvani Gibbard-Satterthwaiteov teorem, koji se odnosi na funkcije društvenog izbora.

[2]Arrow je u [1] zapravo koristio naziv *possibility theorem*, tj. teorem mogućnosti.

Teorem 5.2 (Gibbard-Satterthwaite). *Neka je* $f : \mathcal{T}^n \to K$ *funkcija društvenog izbora koja zadovoljava aksiome* (a_1) *i* (a_4). *Ako je* $k \geq 3$, *onda postoji* $d \in \{1, \ldots, n\}$ *takav da je* $f = f_d$.

Teorem su neovisno dokazali američki filozof Allan Gibbard [5] i ekonomist Mark A. Satterthwaite [9]. U radu [6] dokazana je ekvivalentnost aksioma jake monotonosti (a_2') i otpornosti na strateško glasanje (a_4). Zbog toga iz surjektivnosti (a_1) i jake monotonosti (a_2') također slijedi da je funkcija društvenog izbora diktatorska ako je $k \geq 3$. Direktan dokaz tog teorema i još jedan dokaz Arrowljeva teorema dan je u [8].

Zadaci.

5.1. Pokažite primjerom da Bordina funkcija društvenog blagostanja F_B ne zadovoljava aksiom (A_2') (jaku monotonost).

5.2. Dokažite da Bordina funkcija društvenog izbora f_B zadovoljava aksiom (a_2) (monotonost). Pokažite primjerom da ne zadovoljava aksiom (a_2') (jaku monotonost).

5.3. Dokažite da u slučaju samo dva kandidata ($k = 2$) Bordina funkcija društvenog blagostanja F_B zadovoljava aksiome (A_2') i (A_3), a Bordina funkcija društvenog izbora f_B zadovoljava aksiome (a_2') i (a_4).

5.4. Dokažite da iz aksioma surjektivnosti (a_1) i otpornosti na strateško glasanje (a_4) slijedi jaka monotonost (a_2').

6. Zaključak

U ovom članku osvrnuli smo se na matematičko modeliranje izbornog procesa u okviru teorije društvenog izbora. Najviše pažnje posvetili smo opisivanju izbornog sustava i svojstava koje treba zadovoljavati jezikom relacija i funkcija. Glavni teoremi o nemogućnosti tvrde da ne postoje funkcije društvenog izbora i blagostanja koje zadovoljavaju aksiome demokracije i nisu diktatorske. Postavlja se pitanje o posljedicama tih rezultata po stvarne izbore i donošenje kolektivnih odluka u praksi. Korištena terminologija nameće zaključak da “demokracija nije moguća”, odnosno da nužno vodi u “diktatorstvo”. Autori ovog članka ne slažu se s tom pojednostavljenom interpretacijom.

Po našem mišljenju teoremi nemogućnosti pokazuju da ne treba pretjerivati s formalnim zahtjevima i pravilima. Pojedinačni aksiomi demokracije motivirani su željom da izbori budu pravedni, ali u kombinaciji dovode do paradoksalnog zaključka o diktatorstvu, tj. ne mogu svi biti zadovoljeni. Svjedoci smo da se slične situacije događaju u praksi. U Hrvatskoj se problem nepoštivanja ili "zaobilaženja" zakona i pravila često pokušava riješiti donošenjem novih pravila. Mnogobrojna i pretjerano komplicirana pravila mogu dovesti do paradoksalnih situacija.

Smatramo da je za stvarne izbore iznimno važno da pravila budu jednostavna, razumljiva i lako provediva. Također, važno je da sudionici izbora (kandidati i birači) prihvaćaju mogućnost da ishod bude suprotan od onog koji priželjkuju. Drugim riječima, za uspjeh demokracije potrebna je dovoljno razvijena "demokratska svijest" društva, a ne samo dobra izborna pravila.

7. Rješenja zadataka

2.1. Neka je relacija ρ irefleksivna i tranzitivna. Neka vrijedi $x\,\rho\,y$ i $y\,\rho\,x$. Tada zbog tranzitivnosti slijedi $x\,\rho\,x$, što ne može biti istina zbog irefleksivnosti. Dakle, nikad se ne događa $x\,\rho\,y$ i $y\,\rho\,x$, pa je implikacija da iz $x\,\rho\,y$ i $y\,\rho\,x$ slijedi $x = y$ istinita "na prazno".

2.2. Na skupu $\{1, 2, 3\}$ definiramo relaciju $\{(1, 1), (1, 3), (2, 2), (3, 3)\}$. Ta relacija je parcijalni uređaj u kojem neusporedivost nije tranzitivna: 1 nije usporediv s 2 i 2 nije usporediv s 3, ali 1 jest usporediv s 3.

2.3. Neka je $\prec$ tranzitivna relacija za koju je i neusporedivost tranzitivna te neka vrijedi $x \prec y$. Pretpostavimo da postoji z takav da je $x \not\prec z$ i $z \not\prec y$. Tada vrijedi $z \not\prec x$, jer bi u suprotnom iz tranzitivnosti slijedilo $z \prec y$. Analogno vidimo da vrijedi $y \not\prec z$. Dakle, x i y nisu usporedivi sa z, a međusobno su usporedivi. To je kontradikcija s tranzitivnosti usporedivosti pa takav z ne postoji.

Obrnuto, neka relacija $\prec$ zadovoljava uvjet iz zadatka. Neka je x neusporediv sa z i z neusporediv sa y. Zbog uvjeta ne može vrijediti $x \prec y$ niti $y \prec x$, pa su x i y također neusporedivi i neusporedivost je tranzitivna.

2.4. Neka je $\prec$ strogi slabi uređaj na K. Neusporedivost obzirom na $\prec$ je relacija ekvivalencije na K, tj. relacija $\sim$ koja je refleksivna, simetrična i tranzitivna. Relacija $\prec$ inducira totalni uređaj na skupu klasa ekvivalencije $K/\!\sim$ (stavimo $[x] \prec [y] \iff x \prec y$, gdje su $[x]$ i $[y]$ klase ekvivalencije kojoj pripadaju x i y). Označimo odgovarajuću funkciju korisnosti s $v : K/\!\sim\, \to \{1, \ldots, l\}$ (l je broj klasa ekvivalencije). Možemo je proširiti do funkcije $u : K \to \mathbb{R}$ stavljajući $u(x) = v([x])$. Pokazuje se da za tako definiranu funkciju vrijedi

$x \prec y \iff u(x) < u(y)$.

2.5. Po prethodnom zadatku, relaciju strogog slabog uređaja kod koje neusporedivost ima točno l klasa ekvivalencije možemo identificirati sa surjekcijom $u : K \to \{1, \dots, l\}$. Broj surjekcija s k-članog na l-člani skup je $S(k,l) = \sum_{i=0}^{l} (-1)^i \binom{l}{i} (l-i)^k$ (vidi [3, korolar 5.1.2] ili [7, teorem 2.6.3]). Prema tome, ukupan broj relacija strogog slabog uređaja je $\sum_{l=1}^{k} S(k,l) = \sum_{l=1}^{k} \sum_{i=0}^{l} (-1)^i \binom{l}{i} (l-i)^k$. To je takozvani k-ti uređeni Bellov broj, tj. broj uređenih particija k-članog skupa.

3.1. Neka je $(\prec_1, \dots, \prec_n) \in \mathcal{T}^n$ profil individualnih preferencija i neka su $x_1, x_2 \in K$ dva Condorcetova pobjednika za taj profil. Tada bi po definiciji broj individualnih preferencija u kojima je x_1 bolje rangiran od x_2 trebao biti istovremeno veći i manji od broja individualnih preferencija u kojima je x_2 bolje rangiran od x_1, što je nemoguće. Dakle, ne mogu postojati dva Condorcetova pobjednika.

3.2. Na opisani način možemo dobiti svaku funkciju društvenog izbora $f : \mathcal{T}^n \to K$. Zaista, za zadanu funkciju f definiramo funkciju društvenog blagostanja $F : \mathcal{T}^n \to \mathcal{T}$ tako da $f(\prec_1, \dots, \prec_n)$ bude najbolje rangirani kandidat u društvenoj preferenciji $F(\prec_1, \dots, \prec_n)$, a ostale kandidate poredamo proizvoljno. Tada se funkcija društvenog izbora pridružena funkciji F podudara s f.

4.1. Neka je $P' \in \mathcal{T}^n$ profil za kojeg je kandidat $x \in K$ najbolje rangiran u svakoj od komponenti. Prema surjektivnosti (a_1), postoji profil $P \in \mathcal{T}^n$ takav da je $x = f(P)$. Očito vrijedi $P' \leq_x P$, pa iz aksioma (a_2') slijedi $x = f(P')$. Time je dokazan Paretov uvjet (a_1').

4.2. Iz aksioma (A_1) (surjektivnosti funkcije društvenog blagostanja) očito slijedi aksiom (A_1'). Obrat ne vrijedi – kao primjer možemo uzeti $K = \{x, y, z\}$ i bilo koju funkciju društvenog blagostanja koja za sliku ima sljedeće 3 od 6 mogućih permutacija kandidata: (x, y, z), (y, z, x), (z, x, y). Takva funkcija nije surjekcija, ali zadovoljava aksiom (A_1') jer su mogući svi poretci parova kandidata. Slično, iz aksioma (A_1'') slijedi Arrowljev aksiom (A_1'), a obrat ne vrijedi. Iz aksioma (A_1'') slijedi i surjektivnost (A_1). Zaista, ako vrijedi (A_1''), onda za bilo koji uređaj $\prec \in \mathcal{T}$ vrijedi $F(\prec, \dots, \prec) = \prec$. Nije teško naći primjer surjektivne funkcije društvenog blagostanja koja ne zadovoljava aksiom (A_1'').

Uz dodatnu pretpostavku jake monotonosti (A_2') pokazuje se da iz (A_1') slijedi (A_1''), analogno kao u prethodnom zadatku. Tada su sva tri zahtjeva (A_1), (A_1'), (A_1'') međusobno ekvivalentna.

5.1. Neka na izborima sudjeluju tri kandidata x, y, z i dva birača. Neka su profili $P = (\prec_1, \prec_2)$ i $P = (\prec_1', \prec_2')$ zadani funkcijama korisnosti u_1, u_2 i u_1', u_2':

	x	y	z
u_1	1	3	2
u_2	3	2	1
u	4	5	3

	x	y	z
u_1'	1	2	3
u_2'	3	1	2
u'	4	3	5

Bordine društvene preferencije $\prec = F_B(P)$ i $\prec' = F_B(P')$ odgovaraju funkcijama $u = u_1 + u_2$ i $u' = u_1' + u_2'$. Vidimo da je $x \prec y$ i $P' \leq_x P$, ali ne vrijedi $x \prec' y$. Dakle, funkcija F_B ne zadovoljava aksiom (A_2').

5.2. Neka profilima $P = (\prec_1, \ldots, \prec_n)$ i $P' = (\prec_1', \ldots, \prec_n')$ odgovaraju funkcije korisnosti $u_1, \ldots, u_n$ i $u_1', \ldots, u_n'$. Suma $u = \sum_{i=1}^n u_i$ je minimalna za pobjednika izbora $x = f_B(P)$. Iz $P' <_x P$ slijedi $u_i'(x) \leq u_i(x)$ i $u_i'(y) \geq u_i(y)$, za sve kandidate $y \in K$, $y \neq x$ i birače $i \in \{1, \ldots, n\}$. Prema tome, suma $u'(x) = \sum_{i=1}^n u_i'(x)$ nije se povećala u odnosu na $u(x)$, a $u'(y)$ se nije smanjila u odnosu na $u(y)$, $y \neq x$. Stoga u' također poprima minimalnu vrijednost za kandidata x i vrijedi $x = f_B(P')$. Dakle, Bordina funkcija f_B zadovoljava monotonost (a_2). Iz sljedećeg primjera s četiri kandidata x, y, z, w i dva birača vidi se da ne zadovoljava jaku monotonost (a_2').

	x	y	z	w
u_1	1	3	4	2
u_2	3	2	1	4
u	4	5	5	6

	x	y	z	w
u_1'	1	2	3	4
u_2'	3	1	2	4
u'	4	3	5	8

Vrijedi $x = f_B(P)$ i $P' \leq_x P$, ali $f_B(P')$ nije x nego je y.

5.3. Za $k = 2$ funkcija F_B zadovoljava aksiom (A_3) jer su tada jedini neprazni podskupovi $S \subseteq K$ jednočlani, pa se restrikcije društvenih preferencija sigurno podudaraju. Za aksiome (A_2') i (a_2') ključno je što za $k = 2$ postoje samo dvije permutacije skupa K. U tom slučaju za profile vrijedi $P' \leq_x P$ ako i samo ako vrijedi $P' <_x P$, pa je aksiom (A_2') ekvivalentan s (A_2) i (a_2') je ekvivalentan s (a_2). Znamo da F_B zadovoljava (A_2) i f_B zadovoljava (a_2). Konačno, spomenuli smo da je (a_4) ekvivalentan s (a_2). Zato Bordina funkcija društvenog izbora f_B za $k = 2$ zadovoljava i aksiom (a_4).

5.4. Kratak dokaz te tvrdnje nalazi se u [8].

Literatura

[1] K.J. Arrow, *Social choice and individual values*, Wiley, 1951. (prvo izdanje); 1963. (drugo izdanje). Dostupno na: `http://cowles.econ.yale.edu/P/cm/m12-2/`

[2] M. Botić, *Arrowljev teorem*, diplomski rad, PMF-Matematički odsjek, Sveučilište u Zagrebu, 2011.

[3] P.J. Cameron, *Combinatorics: topics, techniques, algorithms*, Cambridge University Press, 1994.

[4] D. Ciraki, *Teorija javnog izbora i paradoksi glasovanja*, Politička misao **33** (1996), 198–225. Dostupno na: `http://fakultet.fpzg.hr/politicka-misao/DataStorage/Articles/900.pdf`

[5] A. Gibbard, *Manipulation of voting schemes: a general result*, Econometrica **41** (1973), 587–601.

[6] E. Muller, M.A. Satterthwaite, *The equivalence of strong positive association and strategy-proofness*, Journal of Economic Theory **14** (1977), 412–418.

[7] I. Nakić, *Diskretna matematika*, skripta, PMF-Matematički odsjek, Sveučilište u Zagrebu, 2009. Dostupno na: `http://web.math.hr/nastava/komb/predavanja/predavanja.pdf`

[8] P.J. Reny, *Arrow's theorem and the Gibbard-Satterthwaite theorem: a unified approach*, Economics Letters **70** (2001), 99–105. Dostupno na: `http://home.uchicago.edu/~preny/papers/arrow-gibbard-satterthwaite.pdf`

[9] M.A. Satterthwaite, *Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions*, Journal of Economic Theory **10** (1975), 187–217.

[10] Wikipedia, *Borda count*, svibanj 2011. `http://en.wikipedia.org/wiki/Borda_count`

[11] Wikipedia, *Instant-runoff voting*, svibanj 2011. `http://en.wikipedia.org/wiki/Instant-runoff_voting`