

# **Člověk v zrcadle teorie her**

**Jak nám matematika a filozofie pomáhají zjišťovat, co jsme zač**

**Jaroslav Peregrin**

# Obsah

## Úvod

Co má člověk dělat?

Teorie her

Co je v této knize

## Kdo s koho (aneb Kompetitivní hry)

miniNIM

NIM

Šachy

Další příklady her a optimální strategie

## Když on takhle, tak já takhle (aneb Strategie, ekvilibria a optimalita)

Další pojmy teorie her

Nashovo ekvilibrium a optimalita

Hicksovsky a paretoovsky optimální strategie

## Kdo s kým o čem pro koho (aneb Kooperativní hry)

Koordinace

Vězňovo dilema a spolupráce

Zoon politicon

## Komu můžeme věřit (aneb Důvěra)

Stonožka a zpětná indukce

Iterované hry a "lidové teorémy"

Signalizace

## Já já já, jenom já (aneb Člověk člověku vlkem?)

Ultimátní hra, diktátorská hra a společné statky

Homo economicus?

Je člověk nutně sobec?

## Kdo jsme a jak jsme se tím stali (aneb Evoluce)

Jak se z nás stalo to, co jsme

Evoluce a spolupráce

Hry jako možné modely vzniku spolupráce

Dynamický pohled, evolučně stabilní strategie

## Proč děláme nejen to, co musíme, ale i to co máme (aneb Řád)

Reciproční altruismus, altruistické trestání, výběrová spolupráce,...

Skupinový výběr

Člověk a jeho pravidla

## Literatura

# Úvod

## ***Co má člověk dělat?***

*Poznej sám sebe*, nápis nad vchodem do delfské věštírny byl odedávna brán jako výzva filozofům. Co jsme my lidé zač, čím se lišíme od jiných tvorů a bytostí, odkud jsme se vzali a proč děláme to, co děláme? Cest, kterými se myslitelé snažili tuto výzvu naplnit, známe z historie bezpočet. Navíc už to není zdaleka jen filozofie, která nám pomáhá odhalovat naši podstatu. Vědy se v různých specifických ohledech dostávají ve zkoumání nás lidí samozřejmě mnohem hloub. Na filozofii pak zbývá úloha takové dílčí poznatky poskládat do koherentního obrazu.

Některé vědy se soustřeďují pouze na člověka, těm se tradičně říká *humanitní* (v angličtině se jim ovšem vůbec neříká *sciences*, tedy vědy, ale jenom *humanities*). V posledních desetiletích se ale spousta pozoruhodných poznatků odvinula od toho, že jsme se sami na sebe začali důsledněji než dříve dívat prizmatem *přírodních* věd jako na integrální součást přírody. (Když v roce 1967 publikoval Desmond Morris knihu *Nahá opice*, která propagovala právě takový pohled, byl z toho docela poprask; dnes už se to jeví kontroverzní málokomu.) Zejména evoluční biologie (o níž bude řeč v posledních kapitolách) nám dokázala pozoruhodným způsobem objasnit, v čem spočívá naše odlišnost od jiných tvorů a jak jsme se k ní mohli dopracovat. V této knize se ale pokusíme přiblížit, jak k tomuto objasnění přispívá věda, která nemá na první pohled s výklady povahy člověka mnoho společného, totiž matematika. Půjde ovšem o poněkud okrajovou oblast matematiky, starou ani ne sto let a od svých počátků zaměřenou na lidské konání – o teorii her.

Důležitou součástí filozofie bylo od počátku to, čemu se později začalo říkat *praktická filozofie* a co souviselo s otázkami po tom, jak má člověk rozumně žít. Vedle těch týkajících se etiky, tedy otázek, jak se v různých situacích zachovat správně ve smyslu morálky, se to týkalo i přizemnějších otázek po tom, jak to udělat, aby člověk dosáhl cílů, které si stanovuje.<sup>1</sup> Schopnost volit prostředky přiměřené cílům se později začala nazývat *instrumentální racionalitou*. Během posledních desetiletí se tato schopnost stala předmětem studia tzv. teorie racionální volby.

Představme si, že mohu udělat několik různých věcí a stojím před volbou, kterou z těchto možností vybrat. Pokud vím, k jakým důsledkům moje činy povedou, a dovedu je srovnat z hlediska toho, jak jsou pro mě žádoucí, je volba jednoduchá: vykonám to, co povede k výsledku, který je pro mě nejlepší. Čím je dáno to, co je pro mě více a co méně žádoucí? Je to zřejmě věc cílů, jež si stanovuji, a hodnot, které zastávám. Stojím-li například před rozhodnutím, zda si koupit auto či zda peníze raději investovat do svého vzdělání, a rozhodnu se pro druhou možnost, je to zřejmě proto, že vzdělání samo o sobě již považuji za hodnotné nebo jej vidím jako účinný prostředek k dosahování nějakých svých dalších cílů, třeba vydělávání peněz.

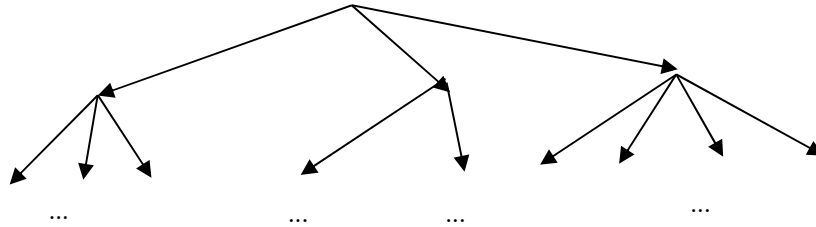
---

<sup>1</sup> Kant třídí pravidla, neboli *imperativy*, kterými se řídíme, na *hypotetické*, které jsou podmíněny specifickými cíli, a *kategorické*, které bychom měli dodržovat bezpodmínečně.



Mám tedy dvě možnosti, jak se do Jedové chýše dostat: buďto mohu jít nejprve doleva a pak dvakrát doprava, nebo naopak nejprve doprava a pak dvakrát doleva. (Která z těchto cest je lepší? To se z daného grafu nedá říci. Kdybychom ovšem měli například u každé šipky uvedenou příslušnou vzdálenost, mohli bychom zjistit, která z nich je kratší.)

Obecně můžeme říci, že rozhodování znamená průchod nějakým „stromem“:



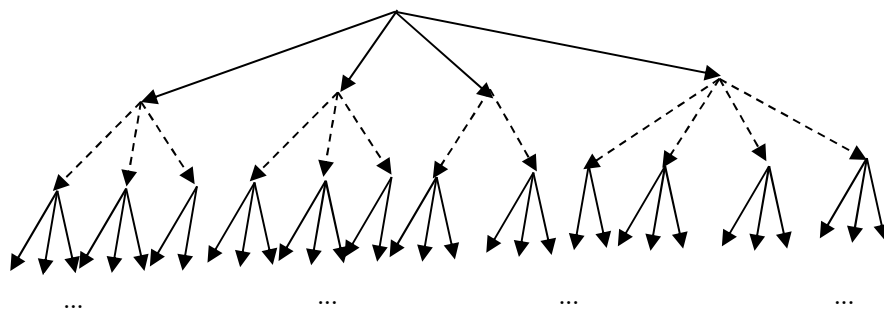
Jak už jsme ale zmínili, může se takováto jednoduchá situace různým způsobem komplikovat (například tím, že výsledky našich akcí budou dány nikoli jednoznačně, ale jenom s určitou pravděpodobností).

Jedna specifická potenciální komplikace přitom stojí za zvláštní pozornost. Může se stát, že se prostředí mění v závislosti na našich rozhodnutích. Není to příliš pravděpodobné v případě cesty do Jedové chýše (že by se ulice v závislosti na mé volbě nějak přeskupily by bylo snad představitelné ve světě Harryho Pottera, ale stěží v tom našem), očekávat to ale můžeme, je-li ono prostředí tvořeno jinými lidmi.

Představme si, že se rozhoduji nikoli o tom, jak se dostat do Jedové chýše, ale o tom, jak pak přesvědčit svého kamaráda, že tam mají lepší pivo než v Zátíší. Tentokrát si nemohu namalovat stejný strom jako v předchozím případě. Jde o to, že rozhodnutí, jakým výrokem otevřu naši diskusi, samo o sobě nevede k určení možností, jež se tím přede mnou otevrou s ohledem na její pokračování. Podstatné totiž bude, jak bude můj přítel na můj výrok reagovat. Řeknu-li například na začátek: „Tady v Jedové chýši čepují především ejly, a nikoli ležáky,“ a připravím-li si pokračování: „A ejly jsou obecně víc hořké než ležáky,“ stává se toto pokračování nepoužitelné, bude-li můj spolustolovník na mou úvodní větu reagovat: „No a já mám raději ležáky, protože nejsou tak hořké.“

Abychom docenili důležitost tohoto druhu rozhodování, musíme uvážit fakt, že člověk je *společenský tvor*. To znamená, že je pro něj stejně tak důležité jako zvládat obcování s přírodou (obstarávat si potravu, vyhýbat se predátorům atd.) I zvládat obcování s jinými lidmi (získat v rámci svého společenství rozumnou pozici, podílet se na rozhodování ap.). V dnešní moderní době centrálně řízených států je dokonce zvládání proplouvání lidskou společností o dost podstatnější než vypořádávání se s přírodou (získávání potravy či vypořádávání se s nebezpečím už obvykle neděláme vlastníma rukama a nohama, vše se odehrává na společenské úrovni a obratnost v jednání s jinými lidmi je tak pro nás podstatnější než síla nebo rychlost).

Strom podobný předchozímu bych si tedy mohl namalovat tak, že bychom každou úroveň mého rozhodnutí proložili úrovní, kdy se rozhoduje můj „protihráč“ či „protihráči“.



A tím se dostáváme k teorii her: ta se nám snaží pomoci s nalezením optimální strategie v případě konfrontací tohoto typu.

### **Teorie her**

Každý, kdo někdy hrál šachy či podobnou hru, se asi zamýšlel nad tím, jak to udělat, aby zvítězil. A možná ho napadlo i to, že by se dalo hrát tak, aby to k vítězství vedlo neodvratně – tedy že by se dala vymyslet nějaká zaručeně vítězná strategie. Pro šachy takovou strategii zatím nikdo nevymyslel (naštěstí! Kdyby ano, byla by to pro tuto hru samozřejmě pohroma), spekulovat o její existenci však jistě můžeme. A rozhodně můžeme spekulovat o tom, jaké strategie jsou lepší a jaké horší.

Teorie her vznikla mimo jiné proto, aby nám do úvah o možných strategiích umožnila zapojit matematiku. A nejde tu ani tak o hry v pravém slova smyslu, jako o reálné situace, v nichž na našich rozhodnutích záleží a které se dají jako všelijaké hry modelovat. Iničiátoři teorie her, matematik John von Neumann a ekonom Oskar Morgenstern, chtěli tuto teorii využít pro studium lidských interakcí a konfliktů zejména ekonomické povahy. V roce 1944 publikují přelomovou knihu *Teorie her a ekonomické chování*,<sup>3</sup> od které se tato teorie odvinula. V úvodu píší:

Důležitost společenských jevů, bohatost a různorodost jejich projevů a složitost jejich struktury je minimálně taková jako u jevů fyzikálních. Dá se tedy očekávat (nebo se obávat), že k tomu, abychom na tomto poli dosáhli rozhodujícího úspěchu, bude třeba matematických objevů významu srovnatelného s matematickou analýzou. (...) A fortiori není pravděpodobné, že bychom si i pro společenské jevy vystačili s opakováním onoho souboru triků, který nám tak dobře posloužil v případě fyziky. To je skutečně velmi nepravděpodobné, protože, jak ukážeme, v našich diskusích narazíme na některé matematické problémy, které jsou zcela jiné než ty, jež se vyskytují ve fyzikální vědě. (S. 6)

Teorie her se skutečně rozrostla do docela košaté disciplíny.<sup>4</sup> Ale čtenář se nemusí obávat. Zde se jí budeme dotýkat jenom po povrchu a nebudeme se pouštět do žádné složité matematiky.

<sup>3</sup> Viz von Neumann & Morgenstern (1944).

<sup>4</sup> Viz např. Osborne (2004), Webb (2007) nebo Harrington (2008). V češtině např. Dlouhý & Fiala (2007) či Kruml (2018); a zejména Binmore (2007). Mnoho materiálů k teorii her lze nalézt i na stránkách <http://www.gametheory.net/>.

Vezměme si příklad, jenž se týká podstatnější věci než vybírání lepší hospody či piva. Představme si, že obyvatelé nějaké obce mají společný projekt, třeba obecní pastvinu. Ta má určitou „kapacitu“ – optimálně užíví tisíc krav. Ve vesnici je sto chovatelů krav, kteří své stádo na obecní pastvinu vyhánějí. Pokud tam každý z nich pošle deset krav, bude „kapacita“ pastviny naplněna a bude se v tomto ohledu jednat o optimální stav.

Fakticky je tu ale pokušení vyhnat na obecní pastvinu nikoli jenom deset, ale jedenáct krav. Bude-li tam tisíc jedna místo tisíce krav, nic dramatického se přece nestane. A ten, kdo tam vyžene svou jedenáctou krávu, z toho bude ve srovnání s ostatními výrazně profitovat. Jestliže to ale udělá jeden, je pravděpodobné, že ostatní nezůstanou pozadu. Budou-li pak všichni na pastvu vyhánět jedenáct krav, někoho napadne vyhnat jich tam dvanáct a tak dále. Konečným výsledkem pak může být to, že krávy pastvinu zcela zaplaví, a tím ji zdevastují. Nakonec tedy nebude mít nikdo nic. Takovému scénáři se v literatuře říká *Tragédie obecní pastviny*.<sup>5</sup>

Nebo si představme jinou situaci. Vesničané potřebují opravit hráz rybníka, který hrozí jejich vesnici zatopit. Každý z nich má eminentní zájem, aby dali dohromady dostatek peněz na postavení hráze. Každý jednotlivě na tom ale bude samozřejmě tím lépe, čím méně bude muset přispívat on osobně. Tím vzniká prostor pro jisté strategické rozhodnutí: z čistě ekonomického hlediska je optimální do projektu investovat tolik, kolik je nezbytně nutné, aby se uskutečnil, ale ani o chlup více. Celá situace může být modelována jako jistý druh jednoduché „hry“, v níž jednotlivé „tahy“ (investice) vedou k různým výsledkům (prostřednictvím určitých „pravidel“, která mohou zachycovat třeba to, jak má příspěvek souviset s majetkem přispívajícího, jak výše mé investice mohou ovlivňovat výše investic ostatních atd.) Optimální strategie pak může být hledána matematickou analýzou příslušné hry.

Tento příklad mimochodem naznačuje, že „hry“, jakými se teorie her zabývá, nemusí připomínat hry, jako jsou šachy – tedy střetnutí dvou hráčů, ve kterém jeden vyhrává a druhý prohrává. Může jít o jiné druhy střetnutí, jichž se účastní více hráčů a kde nemusí platit, že jeden může získat jenom to, co někdo jiný ztratí.

Teorie her si od čtyřicátých let minulého století, kdy byla von Neumannem a Morgensternem iniciována, postupně začala hledat uplatnění i daleko za hranicemi ekonomie, kde vznikla. Proniká do sociologie, politologie i jiných společenských věd a v posledních desetiletích se výrazně zabydlela i mezi přírodovědci zabývajícími se evoluční teorií. Matematici vypracovali komplikované metody, které mohou ti, kdo chtějí pomocí teorie her modelovat reálné jevy, s větším či menším úspěchem využít.

Tažení teorie her napříč vědeckými disciplínami se samozřejmě nemohlo vyhnout ani filozofii. Jednak proto, že sousedí s mnohými disciplínami, do kterých tato teorie pronikla, a pak především i proto, že teorii her bylo možné využít k osvětlení některých skutečně filozofických otázek. Příkladem je jeden z problémů, kterými se budeme dále v knize zabývat, totiž problém *důvěry*.

---

<sup>5</sup> Viz Hardin (1968).

Zdá se, že důvěra je něčím, co je zásadně důležité z hlediska našeho soužití. Jen stěží je možné si představit, že bychom mohli s někým spolupracovat, aniž bychom mu alespoň trochu důvěřovali. Přitom z hlediska čistě instrumentální racionality se důvěra nezdá být tak docela opodstatněná: člověk jistě získá víc, když něčí důvěru zneužije, než když mu ji oplatí. Jistě, instrumentální racionalita není to jediné, čím se v životě řídíme. Dalo by se ale očekávat, že je to minimální základ, k němuž se vyšší patra jenom přidávají, aniž by ho tím narušila.

Tento problém vyvstává v ještě palčivější podobě, když si uvědomíme, že jsme se na světě, stejně tak jako jiní tvorové, objevili v důsledku působení evoluce a přírodního výběru. Jsme tedy potomky těch z našich předků, kteří pro sebe dokázali získat maximum dostupných zdrojů potřebných k tomu, aby přežili a zplodili co nejvíce dětí. A zdá se, že taková úspěšnost je důsledkem maximalizace instrumentální racionality – schopnosti co nejlépe využít prostředků pro své cíle. Je-li tedy důvěra z hlediska instrumentální racionality neopodstatněná, jak se do našeho světa vůbec mohla dostat? Zdá se, že ti, kdo důvěru jiných neopětovali, ale jenom ji zneužívali pro svou potřebu, na tom v tomto ohledu byli lépe než příliš důvěřiví jedinci. Jistě, ti první jednali z našeho pohledu „nemorálně“, proč by ale měla evoluce favorizovat morálku?

### ***O čem se tu čtenář dál dočte***

V následující kapitole si předvedeme, jak aplikovat teorii her na něco, co si člověk běžně pod pojmem *hra* představí (tedy například šachy či „kámen, nůžky, papír“). Podrobněji si rozebereme hru, která je svou složitostí někde mezi těmito dvěma a které se říká NIM, a ukážeme si, jak může v takové hře probíhat hledání vítězné strategie. Tento druh her, kterým budeme říkat *kompetitivní* (hráč v nich může něco získat jenom na úkor ostatních hráčů), je velmi zajímavý z matematického hlediska. Filozofové jej nicméně neshledávají tak podnětným jako hry *kooperativní* a *semikooperativní*, k nimž obrátíme pozornost ve zbylé části knihy.

Kooperativní hry jsou takové, ve kterých může hráč získat jenom tolik, kolik získají i jeho spoluhráči. Semikooperativní hry pak kombinují kooperativní a kompetitivní prvky. U obou typů ztrácí smysl pojmy *vítězství* a *vítězná strategie* – do popředí se namísto nich dostává pojem *ekvilibría*, které vzniká, když se strategie jednotlivých hráčů protnou tak, že pro nikoho z nich není výhodné svou strategii měnit. Stav ekvilibría je zřejmě v jistém smyslu optimální, ale ukážeme si, že můžeme uvažovat i o jiných pojmech optimality.

V další kapitole si předvedeme některé jednoduché typy kooperativních a semikooperativních her a naznačíme, jak lze pomocí nich modelovat a analyzovat některé mezilidské konfrontace, které souvisejí zejména se spoluprací a jejím ustanovováním. Zvláštní pozornost budeme věnovat tzv. Vězňovu dilematu, jež má mezi modely takových konfrontací jisté výsadní postavení.

Dále se budeme zabývat pojmem *důvěry*. Důvěra je něco, co se zdá být nezbytné pro spolupráci a také – alespoň do nějaké míry – rozumné. Avšak analýza pomocí teorie her ukazuje důvěru v poněkud jiném světle: z čistě instrumentálního hlediska vypadá zcela



nerozumně. Zdá se tedy, že důvěra má co dělat s *překonáním* instrumentální racionality, a pozoruhodnou otázkou je, odkud se vlastně bere.

Následně budeme věnovat pozornost hrám, jež bývají využívány k testování toho, zda se člověk normálně opravdu chová podle pravidel instrumentální racionality. Výsledky takových testů jednoznačně ukazují, že tomu tak není a že *Homo sapiens* tedy není (jenom?) *Homo economicus*. To bývá obvykle vysvětlováno tím, že je člověk geneticky vybaven mj. i jistým smyslem pro *fair play*.

Problém „odkud se co u člověka bere“ vyvstává v nových obrysech, když se teorii her pokusíme aplikovat na to, jak mohl náš živočišný druh vzniknout v rámci evoluce. Ta totiž není zodpovědná jenom za tělesné orgány a další fyzické rysy, ale i za všelijaké vzorce chování (případně dispozice k nim). To otevírá možnost zcela nového druhu aplikace teorie her: zjistit, jaký vzorec chování je v rámci přírodního výběru výhodnější, můžeme pomocí analýzy hry, jejímiž strategiemi jsou právě takové vzorce. Pak ale musíme mít na paměti, že už nejde o strategie dvou rozumných hráčů (kteří jsou schopni se od nich odchýlit), ale o souboj tvorů *de facto* ztělesňujících dané strategie (takže nepřipadá v úvahu, že by se chovali jinak).

V tomto kontextu vyvstává problém vzniku spolupráce se zcela novou intenzitou. Jeho rozboru se budeme věnovat v poslední kapitole knihy. Naznačíme cesty, jakými jsme se mohli stát společenskými a spolupracujícími tvory, přestože některé z výše uvedených modelů předpovídají, že by k tomu dojít nemělo.