

Projekt i opracowanie graficzne  
Darek Kondefer

Redakcja  
Katarzyna Bartuzi

© Copyright by Kancelaria Sejmu  
Wydawnictwo Sejmowe Wydanie  
pierwsze, Warszawa 2014

© Copyright by Kazimierz Rzążewski, Wojciech Słomczyński i Karol Życzkowski 2014  
Ilustracje © copyright Marek Wójcikiewicz 2014



# i Każdy głos się liczy !

Wędrownica przez krainę wyborów

Kazimierz Rzążewski • Wojciech Słomczyński • Karol Życzkowski

Ilustrował Marek Wójcikiewicz

Wydawnictwo Sejmowe  
Warszawa 2014

*Surya Bonaly dostała srebro,  
Nicole Bobek – brąz,  
a Kenneth Arrow – Nobla.*

Komentarz matematyka do wyników  
mistrzostw świata w łyżwiarstwie  
figurowym w 1995 roku<sup>1</sup>

1) Warren D. Smith, The Center for Range Voting.

W wielu dziedzinach życia spotykamy się z następującym naturalnym problemem: zespół fachowców ma wybrać jedną osobę z określonej grupy kandydatów. Przykładem niech będzie wybór miss piękności spośród dziewcząt biorących udział w konkursie lub wybór mistrza kulturystyki z grupy wielu dobrze umięśnionych zawodników. Z pozoru zagadnienie wydaje się proste: każdy ze znających się na rzeczy członków jury uważnie przygląda się wszystkim kandydatom, a następnie sekretarz komisji zbiera głosy sędziów i na ich podstawie określa wyniki konkursu lub zawodów. I tu pojawia się problem, który znamy już dobrze z poprzednich rozdziałów: w jaki sposób głosy pojedynczych ekspertów przełożyć na spójny werdykt całego jury wyznaczający zwycięzcę konkursu?



Tegoroczne wybory miss piękności naszej oazy wygrała kandydatka numer trzy.

Zauważmy, że podobny problem dotyczy wielu dyscyplin sportu, sztuki i nauki. W rozmaitych dziedzinach przyjęto zwyczajowo różne zasady postępowania, ale trudno obiektywnie stwierdzić, który system wyłaniania zwycięzcy jest najlepszy. Chciałoby się przyjąć, że wszyscy członkowie jury są kompetentni i działają w dobrej wierze, czyli starają się dokonać oceny obiektywnie. Nie podejmiemy się dokonania systematycznej klasyfikacji wszystkich rozwiązań stosowanych przy wyborach laureatów konkursów i zwycięzców zawodów. Zamiast trzymać się ściśle kolejności podanej w tytule rozdziału, przegląd wybranych metod wyłaniania najlepszych rozpoczniemy od sportu.

## 11.1. Skoki narciarskie, jazda figurowa na łyżwach, zapasy i lekka atletyka

### Noty za styl skoku

Ocena zawodnika biorącego udział w konkursie **skoków narciarskich** składa się z dwóch części: punktów za odległość oraz noty za **styl**. Od sezonu 2009/2010 ocena ta jest korygowana przez dodanie lub odjęcie punktów w zależności od siły wiatru i długości rozbiegu. Ponieważ zarówno elektronicznie mierzona odległość skoku, jak i dodatkowe punkty wyznaczane są bez ingerencji jury, interesować nas będzie głównie ocenianie stylu.

Każdy skok ocenia 5 sędziów, którzy przyznają noty od 0 do 20 punktów, biorąc pod uwagę trzy jego fazy: lot, lądowanie i odjazd po nim. Za popełnienie różnych błędów przez skoczka sędzia odejmuje punkty od maksymalnej noty według jasno określonych reguł: za lot – do 5 punktów, za lądowanie – też do 5, wreszcie za odjazd – do 7. Punkty można stracić na przykład za niewłaściwą sylwetkę podczas lotu, złe prowadzenie nart, niecałkowite wyprostowanie nóg oraz za niespokojny lot, czyli nadmierne sterowanie rękoma. Za brak telemarku przy lądowaniu traci się minimum 2 punkty, za podpórkę – od 4 do 5, a za upadek – 7 punktów. Taki szczegółowy taryfikator sprawia, że oceny zawierają w sobie pewną obiektywną składową. Dlatego też noty za styl przyznawane przez sędziów podczas zawodów wysokiej rangi najczęściej nie różnią się dramatycznie.

Aby zmniejszyć wpływ pojedynczego sędziego na wynik zawodnika, w skokach narciarskich stosuje się prostą zasadę odrzucania not skrajnych. Spośród pięciu ocen za styl przyznanych przez pięciu sędziów odrzuca się dwie skrajne oceny: najwyższą i najniższą, a pozostałe noty sumuje się. W ten sposób skoczek może otrzymać za styl co najwyżej 60 punktów, a na jego wynik nie wpływa błędna lub nieobiektywna ocena pojedynczego sędziego. System ten przypomina więc **głosowanie punktowe**, o którym wspominaliśmy w Rozdziale 7.3.

Maksymalne noty za styl sędziowie przyznają rzadko – pierwszy raz w historii pięć dwudziestek w międzynarodowych zawodach otrzymał austriacki skoczek Toni Innauer podczas konkursu w Oberstdorfie w roku 1976. Częściej natomiast zdarza się, że wszyscy sędziowie oceniają dany skok podobnie lub wręcz tak samo. Jeszcze częściej spotykany jest przypadek, że po odrzuceniu ocen skrajnych trzy uwzględniane noty są takie same lub różnią się zaledwie o pół punktu. Analizując przebieg zawodów najwyższej rangi w ciągu ostatnich lat, można uznać, że stosowany



Skoki narciarskie: w powietrzu Adam Małysz – najlepszy polski skoczek pierwszej dekady XXI wieku

system oceny stylu skoku działa w miarę poprawnie. W przeciwieństwie do innych dyscyplin sportu, gdzie również jury ocenia zawodników, w przypadku skoków narciarskich bardzo rzadko zdarza się, aby wyniki zawodów mistrzowskich w tej dyscyplinie sportu uznawane były za wypaczone poprzez oceny sędziów. Ponieważ zawodnicy i kibice są przyzwyczajeni, że wyniki każdego konkursu skoków zależą od warunków atmosferycznych i chwilowych nawet podmuchów wiatru<sup>2</sup>, nieznaczne niedokładności sędziowania można wyrozumiale traktować jako jeszcze jeden czynnik losowy.

## „Dziesiątka” w gimnastyce sportowej

Jedną z najstarszych niewymiernych dyscyplin sportu jest **gimnastyka**. Uczestnicy zawodów w gimnastyce sportowej wykonują układy ćwiczeń na różnych przyrządach. Kobiety startują w czterech konkurencjach (ćwiczenia wolne, skok, poręczce asymetryczne i równoważnia), mężczyźni zaś – w sześciu (ćwiczenia wolne, skok, poręczce, koń z łękami, kółka i drążek). Klasyfikację zawodników prowadzi się osobno w każdej konkurencji oraz w najbardziej prestiżowym wieloboju gimnastycznym, w którym bierze się pod uwagę sumę punktów uzyskanych na wszystkich przyrządach.

Każdy z sześciu sędziów ocenia poszczególne ćwiczenie gimnastyka w skali od zera do 10 punktów, podobnie jak w skokach narciarskich dwie skrajne oceny są odrzucane, a z pozostałych wyciąga się średnią. Dlatego też najwyższa możliwa ocena odpowiada ćwiczeniu wykonanemu absolutnie idealnie, co w praktyce zdarza się rzadko. Starsi miłośnicy gimnastyki pamiętają pierwszy występ Nadii Comăneci podczas igrzysk olimpijskich w Montrealu w roku 1976. Gdy młodziutka gimnastyczka z Rumunii rewelacyjnie wykonała swoje ćwiczenia na poręczkach, publiczność oczekiwała bardzo wysokiej noty, rzędu 9,8–9,9. Tymczasem ku zaskoczeniu kibiców, po momencie oczekiwania na werdykt sędziów, zainstalowana aparatura pokazała nieprawdopodobny wynik 1,0. Dopiero po chwili konsternacji zarówno zgromadzeni widzowie, jak i komentatorzy telewizyjni, odgadli, że jest to błąd systemu prezentacji wyników, który w ogóle nie był zaprojektowany do wyświetlenia liczby **10,0** jednomyślnie przyznanej Rumunce przez wszystkich sędziów!

Ta konkretna historia dotyczy jedynie szczegółu technicznego, ale dobrze ilustruje kluczowy problem, przed jakim stawało ówczesne jury: bardzo wysoko oceniając dokonania wcześniej startujących zawodniczek i chcąc następnie oddać sprawiedliwość wspaniałej Rumunce, sędziowie musieli postawić jej maksymalną notę. Dlatego gdyby w tamtym konkursie któraś z konkurentek występujących po Nadii Comăneci ćwiczyła jeszcze lepiej, sędziowie nie mogliby oczywiście przyznać jej wyższej oceny. Problem stanowiła więc właściwa kalibracja not, które



Gimnastyka sportowa: Nadia Comăneci po wykonaniu ćwiczeń na poręczkach podczas igrzysk w Montrealu na tle kuriozalnej oceny

2) Wpływ zmieniającego się wiatru uwzględnia jednak obecnie wspomniana poprawka, wyliczana za pomocą odpowiedniego algorytmu. Podobna poprawka pozwala porównywać skoki oddane z różnych belek startowych.

każdy z sędziów wystawiał całej grupie startujących gimnastyczek. A w odróżnieniu od skoków narciarskich, w gimnastyce oceny sędziów decydowały i nadal decydują o wszystkim.

Już w latach 90. XX wieku dokonano zmiany systemu sędziowania w zawodach gimnastycznych. Sędziowie podzieleni są teraz na dwie grupy: poza szóstką sędziów oceniających **wykonanie** ćwiczenia (nota *E* jak *execution*) i próbujących wychwycić wszystkie możliwe błędy, dwóch dodatkowych sędziów ocenia **trudność** wykonywanych ćwiczeń i ustala wartość całego układu (nota *D* jak *difficulty*). Do średniej oceny *D* za wartość układu dodaje się średnią z czterech not *E* za wykonanie ćwiczenia, gdyż nadal dwie skrajne oceny z sześciu zostają odrzucone. O ile maksymalna ocena *E* równa się 10,0, to sumaryczna ocena może przyjmować teoretycznie dowolnie duże wartości, jako że górnej granicy oceny *D* za wartość programu nie określono. Takie podejście jest o tyle zasadne, że obecnie gimnastycy wykonują znacznie trudniejsze ćwiczenia niż przed laty.

Idea otwartej skali wartości przypomina nieco skale stosowane w alpinizmie do oceny trudności wspinaczki. W pionierskiej epoce eksploracji Tatr po I wojnie światowej najtrudniejsze ówczesne drogi wspinaczkowe określano mianem „nadzwyczaj trudnych” (tzw. *aussery*<sup>3</sup>) i opisywano stopniem trudności VI. W latach późniejszych życie pokazało, że wspinaczka jest możliwa po drogach znacznie, ale to znacznie trudniejszych od „nadzwyczaj trudnych” dróg szóstkowych. W tej sytuacji postanowiono używać skali otwartej i już bez podawania nazw wprowadzać dalsze stopnie trudności. Obecnie wspinacze przechodzą drogi o trudności VII, VIII, IX, X, a także XI<sup>4</sup>, i sądzić można, że proces otwierania nowych stopni trudności na tym się nie skończy.

Podobnie należy założyć, że stopień trudności wykonywanych ćwiczeń gimnastycznych będzie się z czasem nadal zwiększał. W tej sytuacji ocena *D* za wartość ćwiczeń mogłaby kiedyś przekroczyć 10,5 lub nawet 11,0. Aby jednak w przybliżeniu zachować skalę używanych not, Międzynarodowa Federacja Gimnastyczna wprowadziła zasadę ustalania **bazowej oceny SV (Start Value)**, która wynosiła dla kobiet odpowiednio 9,4 w roku 1994, 9,0 – w roku 1998, a 8,8 – w roku 2002. Dopiero do tej wartości bazowej, dla mężczyzn niższej o 0,4, dodawało się punkty obrazujące na ile prezentowane ćwiczenia są trudniejsze od aktualnego standardu trudności. Powolne obniżanie współczynnika SV zapobiegło szybkiemu wzrostowi średniej przyznawanych ocen.

Po kolejnej zmianie systemu punktowania, od roku 2006 dwóch sędziów oceniających trudność programu korzysta ze szczegółowego taryfikatora (*code of points*). Trudne programy oceniane są teraz na 6–7 punktów, co przy prawie optymalnym wykonaniu prowadzi do całkowitej oceny przekraczającej 16 punktów. I tak na przykład na igrzyskach

3) Z niem.: *ausserordentlich schwierig* – niezwykle trudne.

4) W innej skali stosowanej przez wspinających się w polskich skałkach są to drogi o trudności VI.1, VI.2, aż po VI.8.

olimpijskich w Pekinie w roku 2008 występujący w finale skoku Leszek Blanik w obu próbach popełnił niewielkie błędy przy lądowaniu, ale ponieważ układy, które wykonywał, były bardzo trudne, otrzymał za nie noty 16,600 i 16,475. Średnia z tych dwóch ocen, 16,537, zapewniła Blanikowi złoty medal olimpijski, pierwszy w historii polskiej gimnastyki.

W środowisku czołowych zawodników nie milkną głosy krytyki wobec stosowanego systemu punktacji zawodów gimnastycznych. Często podnosi się argument, że nowy system premiuje stałe podwyższanie trudności ewolucji kosztem jakości wykonania ćwiczeń i uwzględniania walorów artystycznych. Przykładowo, włoska gimnastyczka Vanessa Ferrari w roku 2006 w sposób kontrowersyjny zdobyła tytuł mistrzyni świata w wieloboju gimnastycznym pomimo upadku na równoważni. Punkty stracone w wyniku tego błędu zostały zrównoważone bardzo wysoką oceną poziomu trudności jej programu w ćwiczeniach wolnych. Inni krytycy zmian twierdzą, że obecny system jest mniej zrozumiały dla publiczności, gdyż znosi ocenę 10,0 jako kanon gimnastyki doskonałej.

Z drugiej strony wielu ekspertów przyznaje, że niedawne zmiany systemu idą w dobrym kierunku: sędziowanie jest bardziej obiektywne, a możliwości nadużyć stają się ograniczone. Co więcej, dzięki postępowi technologicznemu sędziowie mogą na bieżąco oglądać powtórzenia ćwiczeń zarejestrowanych przez kamery i w ten sposób jury może rozwiązywać kwestie sporne, opierając się na nagranych materiałach zdjęciowych. Ale gorąca debata nad systemami sędziowania w środowisku gimnastycznym trwa nadal. Kolejne zmiany, w zamierzeniu doskonalące system sędziowania, utrudniają przygotowania sportowców. Nie jest jasne, czy w najbliższych zawodach bardziej będzie się liczyła trudność ćwiczeń, czy ich perfekcyjne wykonanie.

## Skomplikowane reguły i losowanie sędziów w łyżwiarstwie

Finałowe konkursy olimpijskie w łyżwiarstwie figurowym od wielu lat cieszą się wielkim zainteresowaniem publiczności. Do konkurencji olimpijskich należy jazda indywidualna kobiet i mężczyzn, pary sportowe i pary taneczne. W tych ostatnich pary mieszane wykonują tańce na lodzie. Na zawody składa się taniec oryginalny, w którym określono rodzaj układu i charakter muzyki, oraz taniec dowolny<sup>5</sup>, gdzie każda para może dowolnie wybierać temat, muzykę i choreografię. Podczas programu łyżwiarze nie mogą oddalać się od siebie na podwojoną długość ramion. W przeciwieństwie do konkurencji par sportowych, pary taneczne nie wykonują skoków z wieloma obrotami, tancerze nie wyrzucają partnerek w powietrze, a wykonując unoszenie partnerki, tancerz nie podnosi rąk powyżej głowy. W konkurencji par sportowych



<sup>5</sup>) Do roku 2011 wykonywano ponadto taniec obowiązkowy.



Na olimpijskiej tafli brytyjska para Jayne Torvill–Christopher Dean

zawodnicy prezentują dwa układy: program krótki i program dowolny. Poza standardowymi elementami indywidualnymi łyżwiarze wykonują szereg elementów wspólnych: podnoszenia, piruety parami i skoki wyrzucane. Zawody solistów także zaczynają się od programu krótkiego, w którym należy wykonać zadane rodzaje ewolucji. Najlepsi zawodnicy walczą o medale, prezentując dłuższy program dowolny.

Tradycyjny system punktowania występów łyżwiarzy nosi nazwę **systemu szóstkowego**. Nazwa pochodzi od oceny **6,0**, która była dawniej maksymalną, jaką mógł otrzymać zawodnik. Ten „stary system” punktowania został przyjęty na początku XX wieku, gdy zasadniczą rolę w jeździe figurowej odgrywały figury obowiązkowe. Każdy łyżwiarz wykonywał w konkursie 6 różnych figur ocenianych w skali od 0 do 1, co sprawiło że maksymalna ocena każdego występu mieściła się w przedziale od 0 do 6,0. Oglądanie figur obowiązkowych nie było bardzo widowiskowe, więc począwszy od roku 1968, w miarę jak wzrastało znaczenie telewizji, względna waga przykładana do jazdy obowiązkowej stopniowo malała. W roku 1990 figury obowiązkowe w ogóle wypadły z programu zawodów międzynarodowych, ale skala szóstkowa pozostała.

W systemie szóstkowym każdy z dziewięciu sędziów przyznawał za każdy program dwie oceny: jedną za **wartość techniczną** programu, a drugą – za **wrażenie artystyczne**. Obie oceny w skali od zera do sześciu były sumowane z wagami: 40% za program krótki, a 60% za program dowolny. Każdy sędzia stawiał najpierw notę za program, jaka należałaby się zawodnikowi, gdyby program był wykonany bezbłędnie, a następnie odejmował punkty za dostrzeżone błędy. Na podstawie przyznanych ocen każdy z sędziów ustawiał swoją kolejność całej stawki. Aby ustalić, którą lokatę zajął zawodnik w zawodach, ważne były tylko miejsca, na jakich widzieli jego oraz jego konkurentów poszczególni sędziowie, nie zaś przyznane oceny.

Najwyższą ocenę przyznawano niezwykle rzadko. Do historii tej dyscypliny przeszedł taniec do utworu *Boléro* skomponowanego przez Maurice'a Ravela wykonany podczas igrzysk w roku 1984 w Sarajewie przez brytyjską parę Jayne Torvill–Christopher Dean, która za wrażenie artystyczne od wszystkich dziewięciu sędziów otrzymała maksymalną notę 6,0. Na ogół oceny sędziów nie były jednak jednomyślne, a rozbieżność w poglądach jury często miała kluczowy wpływ na końcowe wyniki zawodów.

Obowiązujący pod koniec XX wieku system oceniania zawodów łyżwiarskich nosił nazwę **BOM** (*best of majority*, czyli najlepsza z większości). Dla każdego zawodnika obliczano jego współczynnik **LMR** (*lowest majority rank* – najniższa lokata większości), równy najmniejszej wartości lokaty, powyżej której lub na równi z którą większość (czyli przynajmniej

5 z 9) sędziów plasowała zawodnika. Przykładowo, jeżeli u dziewięciu sędziów zawodnik zajął miejsca: 3, 1, 2, 1, 3, 2, 3, 1, 1, to jego *LMR* wynosiło 2, gdyż na pierwszym lub drugim miejscu sklasyfikowało go sześciu sędziów. Kolejność zawodników ustalano według współczynników *LMR* od najmniejszego do największego. Gdy dwaj zawodnicy mieli takie same *LMR*, obliczano dla nich kolejny współczynnik *SLM* (*size of lowest majority* – rozmiar najniższej większości), czyli liczbę sędziów tworzącą tę większość. Dla naszego przykładowego łyżwiarza *SLM* wynosiło 6. Teraz ten z zawodników, który miał wyższe *SLM*, plasował się wyżej. Gdy i ten współczynnik był dla dwóch łyżwiarzy taki sam, sumowano (jak w systemie Bordy) miejsca i tego z zawodników, który miał mniejszą sumę, klasyfikowano wyżej. Gdy i te sumy się pokrywały, zawodnicy zajmowali *ex aequo* to samo miejsce. Po przeczytaniu tego wszystkiego nikt z czytelników nie będzie zdziwiony, że systemu BOM nie rozumiała ani publiczność, ani sami zawodnicy.

Jak system ten działał w praktyce? Podczas **igrzysk olimpijskich w Lillehammer** (Norwegia) w roku 1994 faworytkami w konkurencji solistek były Amerykanka Nancy Kerrigan i Ukrainka Oksana Bajuł. Po programie krótkim Kerrigan nieznacznie wyprzedzała Bajuł, a w przeddzień decydującego programu dowolnego Ukrainka miała kolizję podczas treningu i doznała kontuzji. W programie dowolnym Amerykanka pojechała świetnie i otrzymała noty 5,7–5,8 za wartość techniczną oraz 5,8–5,9 za wrażenie artystyczne, co dało jej sumaryczną notę 51,9. Startująca później Ukrainka wspaniale wykonała bardzo wymagający program, ale jej lądowanie po potrójnym flipie nie było stabilne. Może dlatego za wartość techniczną programu sędziowie przyznali noty od 5,5 aż do 5,9, a taki rozrzut opinii spotkał się później z dezaprobatą ekspertów. Przy bardziej jednomyślnych ocenach 5,8–5,9 za wrażenie artystyczne całkowita ocena Ukrainki wyniosła 51,7, jedynie o dwie dziesiąte punktu gorzej od Amerykanki. Także po odrzuceniu dwóch skrajnych not suma punktów Bajuł wynosiłaby 40,3 i byłaby niższa od sumy 40,4 uzyskanej przez Kerrigan. Wydawać by się więc mogło, że to Kerrigan powinna zdobyć złoty medal olimpijski. Ale przecież nie oceny były tu ważne, ale kolejność, według której zawodniczki były klasyfikowane przez poszczególnych sędziów. Ponieważ większość, to jest pięciu z dziewięciu sędziów, wyżej oceniła występ Bajuł, podczas gdy jedynie czterech z nich wyżej punktowało Kerrigan, złoty medal zdobyła Ukrainka, a srebrny przypadł Amerykance. Medal brązowy zdobyła Chinka Lu Chen.

Na tym kończy się relacja z lodowiska w Lillehammer. Przypatrzmy się jednak dokładniej punktacji poszczególnych sędziów. Pięciu sędziów punktowało (1) Bajuł, (2) Kerrigan, (3) Chen, następnych dwóch typowało kolejność (1) Kerrigan, (2) Bajuł, (3) Chen, podczas gdy dwóch

pozostałych oceniło łyżwiarki w kolejności (1) Kerrigan, (2) Chen, (3) Bajuł. A więc wszyscy sędziowie widzieli Amerykankę na pierwszym lub drugim miejscu w konkursie, podczas gdy dwóch sędziów uznało, że Bajuł należał się brązowy medal. Gdyby w rachubach i porównaniach uwzględniać też Chen i sumować lokaty uzyskane przez każdą zawodniczkę, to Kerrigan z wynikiem  $5 \times 2 + 2 + 2 = 14$  wyprzedziłaby Bajuł, dla której suma zajętych miejsc wyniosła  $5 + 2 \times 2 + 2 \times 3 = 15$ . Ale do sumowania nie doszło, gdyż współczynnik *LMR* dla Bajuł wynosił 1, dla Kerrigan zaś 2.

Rok po igrzyskach w Lillehammer łyżwiarki spotkały się na mistrzostwach świata. Po przejeździe w programie dowolnym wszystkich czołowych zawodniczek pierwsze miejsce zajmowała Lu Chen, drugie – Amerykanka Nicole Bobek, a trzecie – Francuzka Surya Bonaly. Na taflę wjechała ostatnia łyżwiarka 14-letnia Amerykanka Michelle Kwan, która nie miała już szans na medal. Jej rewelacyjny przejazd sprawił jednak, że Bobek spadła na miejsce trzecie, a Bonaly wskoczyła na drugie, mimo że ich oceny u wszystkich sędziów pozostały, rzecz jasna, takie jak przedtem. Ujawniła się w ten sposób najgorsza wada systemu BOM, jaką opisaliśmy wcześniej: zależność od alternatyw niezwiązanych. Tego typu „przerzutki” (inną opisywaliśmy w Rozdziale 7.3) doprowadzały do białej gorączki wszystkich miłośników łyżwiarstwa i władze **Międzynarodowej Unii Łyżwiarzkiej (ISU)**.

W roku 1998 zastąpiono więc system BOM nowym systemem **OBO** (*one-by-one*, czyli *każdy z każdym*) typu condorcetowskiego, gdzie każdą parę zawodników porównywano między sobą, patrząc, kto ma najwięcej zwycięstw w takich pojedynkach, a w razie remisu stosowano metodę Bordy. Niestety ten system okazał się równie mało zrozumiały dla stronniych jak poprzedni. Jego autorzy wierzyli (nie znając najwyraźniej twierdzenia Arrowa, por. Rozdział 7.3), że uda się im zlikwidować „przerzutki”. Oczywiście tak się nie stało, bo stać się nie mogło.

Kontrowersje w sprawie złota w zawodach solistek na olimpiadzie w Lillehammer zbladły jednak w porównaniu z ewidentnym skandalem sędziowskim, do którego doszło podczas konkursu par sportowych na **igrzyskach olimpijskich w Salt Lake City** w roku 2002. Po programie krótkim prowadziła para rosyjska, Jelena Bierieznaja i Anton Sicharulidze, przed parą kanadyjską Jame Salé–David Pelletier. W programie dowolnym Rosjanie popełnili ewidentny błąd przy podwójnym skoku, podczas gdy para z Kanady swój program wykonała bezbłędnie. Przy ogromnym aplauzie i gromkich okrzykach publiczności: „Sześć, sześć, sześć”, Kanadyjczycy oczekiwali na końcowy wynik, a wielu komentatorów prowadzonych na żywo transmisji telewizyjnych uznało ich za zwycięzców zawodów. Wkrótce okazało się, że otrzymali oni trzy razy

ocenę 5,9 za wartość techniczną programu oraz cztery razy 5,9 za wrażenie artystyczne, a pozostałe oceny były równe 5,8. Nie wystarczyło to jednak do zdobycia złotego medalu, gdyż pięciu sędziów oceniło wyżej parę rosyjską, a jedynie czterech typowało do złota Kanadyjczyków. Pierwsze miejsce dla pary Bieriżnaja–Sicharulidze zaskoczyło komentatorów telewizyjnych i wyraźnie rozżłościło zebraną wokół lodowiska publiczność, która zareagowała głośną dezaprobatą.

Natychmiast po ujawnieniu szczegółowych wyników zwrócono uwagę na pewne korelacje geopolityczne pomiędzy ocenami przyznanymi przez sędziów i zaczęto głośno podważać bezstronność jury. Zanim sędzia z Francji, pani Marie-Reine Le Gougne, dotarła po zawodach do swego hotelu, została zagadnięta przez Angielkę Sally Stapleford, szefa Komitetu Technicznego ISU. Podczas pełnej emocji wymiany zdań Francuzka przyznała, że prezydent Francuskiego Związku Łyżwiarskiego naciskał na nią, aby głosowała na parę z Rosji.

Taka informacja mogła być prawdopodobna zwłaszcza wobec zbliżającego się startu pary francuskiej w finale konkursu par tanecznych, gdzie przychylność sędziego z Rosji mogła się okazać bezcenna. Mimo że sędzia Le Gougne oficjalnie zaprzeczyła później wszelkim naciskom i w pisemnym oświadczeniu stwierdziła, że występ łyżwiarzy rosyjskich oceniła obiektywnie jako najlepszy, sprawą zajął się Międzynarodowy Komitet Olimpijski. Już dwa dni później podjęto bezprecedensową decyzję: zmieniono wyniki zawodów i drugie złoto przyznano parze kanadyjskiej. Formalnie unieważniono oceny sędzi z Francji, co przy czwórce jurorów uznających wyższość pary z Kanady i czwórce wskazujących na Rosjan uzasadniało konieczność wręczenia dwóch złotych medali.

Po wydarzeniach na olimpiadzie w Salt Lake City ISU przystąpiła do opracowania nowych zasad oceny zawodników. Od roku 2004 obowiązuje nowy system punktacji zawodów łyżwiarskich znany pod nazwą **IJS** (*International Judging System*). Po występie każdego zawodnika członkowie trzyosobowego panelu technicznego oceniają stopień trudności poszczególnych elementów, podczas gdy każdy z dziewięciu sędziów ocenia jakość ich wykonania i do oceny bazowej ustalonej przez panel dodaje lub odejmuje od niej do trzech punktów. Po wylosowaniu siedmiu ocen i odrzuceniu z nich dwóch skrajnych wylicza się średnią ocenę za wartość techniczną elementu. Noty te zsumowane z odpowiednimi wagami składają się na łączną ocenę wartości technicznej programu. Do tej oceny dolicza się punkty za składniki (komponenty), w skali od 0 do 10, w ramach których ocenia się połączenia elementów, choreografię i interpretację. Po naliczeniu ewentualnych punktów karnych przydzielanych według precyzyjnego taryfikatora za upadki, przerwy w programie, zbyt długi lub zbyt krótki program oraz niewłaściwy strój otrzymuje

się całkowitą ocenę. Łączne noty ze wszystkich występów sumuje się, a zawody wygrywa łyżwiarz z największą liczbą punktów. W przypadku remisu decyduje punktacja za program wykonany jako ostatni.

Takie reguły wyglądają rozsądnie i sprawiają, że kolejność ustalona przez każdego z sędziów nie ma bezpośredniego wpływu na wynik zawodów. Z drugiej strony obserwatorzy twierdzą, że teoretycznie udoskonalone zasady w praktyce nie okazują się skuteczne, a sposób sędziowania nadal pozostawia wiele do życzenia. Przykładowo, obie wystawiane oceny okazują się silnie skorelowane, więc zawodnik dobrze oceniony za wartość techniczną programu prawie na pewno otrzyma też wysoką notę za składniki. Dlatego dla każdego łyżwiarza kluczowe znaczenie ma kolejność startu w programie krótkim. Nawet najlepsze wykonanie trudnego programu przez zawodnika startującego w grupie słabszych nie pozwoli mu zająć dobrego miejsca w zawodach, gdyż sędziowie nie traktują łyżwiarzy z tej grupy jako potencjalnych zwycięzców.

Aby zmniejszyć możliwe układy i powiązania w gronie sędziów postanowiono dodatkowo, że w zawodach najwyższej rangi sędziować będzie 12 jurorów, ale do końcowej noty uwzględniać się będzie jedynie 9 losowo wybranych ocen, z których dwie skrajne będą odrzucane. W ten sposób żaden sędzia nie mógł być pewien, że jego głos w ogóle będzie się liczył, co w zamysle twórców systemu miało ograniczyć możliwość nieuczciwego wpływania na wyniki zawodów. W roku 2008 zmniejszono liczbę sędziów do 9, spośród ocen których losuje się 7.

Po ogłoszeniu szczegółów nowego systemu powszechnie chwalili go liczni komentatorzy oraz byli zawodnicy najwyższej klasy. Tymczasem do analizy wprowadzonych reguł wzięli się matematycy. Profesor statystyki z Uniwersytetu Yale, **John Emerson**, przebadął wyniki konkurencji solistek na rozegranych we Francji w roku 2006 mistrzostwach Europy. Zawody wygrała Irina Słuckaja z 66,43 punktami, podczas gdy kolejne cztery miejsca zajmowały zawodniczki z dorobkiem 60,88, 60,87, 60,19 oraz 60,04 punktu. Znając oceny przyznane przez wszystkich 12 sędziów, matematyk rozważał, jak wyglądałyby wyniki zawodów przy innym wyborze dziewięćki sędziów, których oceny zostały wzięte pod uwagę. Wyniki takiej analizy były dość zaskakujące. O ile pozycja złotej medalistki, Iriny Słuckiej, nie ulegała wątpliwości, to różnice pomiędzy pozostałymi łyżwiarkami były tak małe, że srebrny medal mogła zdobyć każda z pozostałej czwórki, w zależności od dokonanego losowo wyboru sędziów. A więc czy o przydziale medali na zawodach tej klasy ma rozstrzygać rzut kością do gry<sup>6</sup>? Nie każdemu los sprzyja: mistrzostwa we Francji były pechowe dla Elene Gedewaniszwili. Ostatecznie zajęła ona czwarte miejsce, mimo że w większości możliwych układów losowania sędziów była trzecia.



6) W niektórych dyscyplinach olimpijskich czynnik losowy ma istotny wpływ na wynik. Na przykład w zawodach pięcioboju nowoczesnego przed konkursem skoków przez przeszkody zawodnicy losują konie. Jak można przeczytać w Rozdziale 1.3, jest to stara tradycja, lecz uczestnicy tych zawodów losują jedynie dosiadanego wierzchowca, a nie uzyskane czasy przejazdu czy liczbę zrutek...

Czy istnieje więc jakieś rozsądne rozwiązanie tego problemu? Emerson sugerował, że w omawianym przypadku bardzo równych wyników cała czwórka zawodniczek powinna otrzymać srebrny medal, aby o wynikach zawodów nie decydował ślepy los. To natychmiast rodzi pytanie: jakie wyniki są „bardzo równe”. Jak wyznaczyć przedział, w którym oceny uznajemy za równorzędne?

Inna możliwość redukcji czynnika losowego polega na przeprowadzeniu swoistej eliminacji w gronie... sędziów! W procesie oceny pierwszego programu łyżwiarzy (lub pierwszej kolejki skoków na nartach) uczestniczyliby na przykład komplet 12 sędziów, a do oceny zawodnika wchodziłoby tylko część ocen z odrzuceniem skrajnych. W drugiej części zawodów uczestniczyliby tylko sędziowie z najmniejszą liczbą odrzuconych werdyktów. Tu także odrzucano by skrajne wyniki. Żaden system nie może całkowicie uniemożliwić sędziom przyznawania nieuczciwych ocen, lecz w tej metodzie głosy gorszych lub mniej obiektywnych sędziów w mniejszym stopniu wpływałyby na wynik końcowy<sup>7</sup>.

### Zapasy: zwycięzca wszystkich walk bez olimpijskiego medalu

Olimpijskie zawody w **boksie** organizowane są systemem pucharowym: zawodnik przegrywający walkę odpada z turnieju. W naturalny sposób zwycięzcą zawodów zostaje bokser, który wygrał wszystkie stoczone pojedynki. Przy (nie zawsze zasadnym) założeniu, że każda walka sędziowana jest obiektywnie, złoty medal przypadnie na pewno najlepszemu zawodnikowi igrzysk w danej kategorii wagowej. Już jednak srebrnego medalu nie musi otrzymać drugi w kolejności bokser, który mógł przegrać z późniejszym zwycięzcą we wczesnej fazie turnieju. Aby zminimalizować prawdopodobieństwo takiego zdarzenia, w turniejach tenisowych rozstawia się najlepszych potencjalnie tenisistów i tenisistki, aby zbyt wcześnie nie spotkali się w bezpośrednim starciu.

W innych dyscyplinach olimpijskich obowiązują bardziej skomplikowane reguły, które mogą prowadzić do nieoczekiwanych rezultatów. Podczas **igrzysk olimpijskich w Monachium** w roku 1972 Christo Trajkow z Bułgarii startował w turnieju **zapasów** w stylu klasycznym w wadze do 57 kilogramów. Zapaśnik bułgarski stoczył sześć walk, z których wygrał wszystkie, a mimo to nie zdobył żadnego medalu i ostatecznie zajął dopiero piąte miejsce. Co ciekawsze, w bezpośrednich pojedynkach Trajkow zwyciężył zarówno późniejszego złotego medalistę, Rustama Kazakowa z ZSRR, jak i Hansa-Juergena Veila z RFN, który zdobył srebrny medal. Jak w ogóle były możliwe takie wyniki? Czy była to zwyczajna pomyłka sędziów, czy też może wynik antybułgarskiego spisku?

<sup>7</sup>) Pomimo wszystkich zmian jakie w ostatnich latach zaszły w systemie sędziowania zawodów łyżwiarskich, w czasie **igrzysk olimpijskich w Soczi** w roku 2014 znów pojawiły się oskarżenia o zмовę sędziowską. Tym razem sojusz sędziów rosyjskich i amerykańskich miał jakoby pozbawić szans na pierwsze miejsce kanadyjską parę taneczną.

## Powtórka z krystianii i egzamin instruktorski PZN

Podczas zawodów sportowych rzadko zdarza się, aby z istotnego powodu przerwano rozgrywaną konkurencję, a zawodnikom, którzy już wystąpili, nakazano powtórzenie występu. Po ostatnich zmianach regulaminu nie dochodzi do takich sytuacji nawet w konkursach skoków narciarskich. W mniej prestiżowych zawodach taka sytuacja może się jednak zdarzyć, także na skutek problemów komisji sędziowskiej przy ocenianiu zawodników.



W czasie egzaminu na stopień instruktora Polskiego Związku Narciarskiego (PZN) w Szczyrku, w lutym 1991 roku kandydaci w ostatnim punktowanym przejeździe mieli zademonstrować krystianię nożycową<sup>8</sup>. Po wykonaniu tej ewolucji przez czterech pierwszych narciarzy okazało się, że członkowie Komisji Egzaminacyjnej sami nie są w stanie uzgodnić, jaki sposób odstawiania narty górnej przy skrucie należy uznać za prawidłowy. Aby rozwiązać problem, poproszono Zbigniewa Stanisławskiego (na zdjęciu) – członka Komisji, jednego z najlepszych instruktorów i wykładowców PZN – o demonstrację wymaganej krystianii. Jego bezbłędny przejazd potraktowano jako wzorzec wykonania ewolucji na maksymalną ocenę 10 punktów. Następnie, ku zadowoleniu oczekujących, wznowiono egzamin, a czterej pierwsi kandydaci otrzymali przywilej powtórzenia punktowanego przejazdu.



<sup>8</sup>) Po wprowadzeniu nart taliowanych ta ewolucja przestała mieć znaczenie i wypadła z programu nauczania PZN.



Odpowiedź na to pytanie przynosi analiza ówczesnego systemu rozgrywania turnieju metodą **najmniejszej liczby punktów karnych**. Pojedynek zapaśników kończy się przed czasem, gdy jeden z zawodników rzuci przeciwnika na matę lub położy go na łopatki. Po upływie standardowego czasu walki o jej wyniku decydują punkty przyznawane przez sędziów za wykonywane akcje techniczne. Aby premiować zawodników wygrywających przed czasem, po takim zwycięstwie zawodnik zachowywał czyste konto, podczas gdy za wygraną na punkty otrzymywał 1 punkt karny. Za remis każdy zapaśnik dostawał po 2 punkty karne, za porażkę na punkty uzyskiwał 3, a za porażkę przed czasem – 4. Punkty karne sumowano po każdej turze walk, a zawodnik, który zgromadził 6 lub więcej punktów karnych, odpadał z turnieju.



Taki system wygląda na pozór rozsądnie i stosowany w wielu turniejach we wszystkich kategoriach wagowych i obu stylach walki pozwalał zazwyczaj na wyłonienie bezspornie najlepszego zapaśnika w danej konkurencji. Tymczasem podczas turnieju w Monachium doszło do nietypowej sytuacji: Trajkow wygrywał wszystkie swoje walki na punkty, więc po sześciu turach zgromadził 6 punktów karnych i odpadł z turnieju. Natomiast czterej inni zapaśnicy zwyciężali w swoich pojedynkach przed czasem i mimo jednej porażki na punkty z Trajkowem, kosztującej 3 punkty, z mniejszą liczbą punktów karnych awansowali do kolejnej tury i pomiędzy sobą rozegrali decydujące walki o olimpijskie medale. Trajkow byłby prawdopodobnie został zwycięzcą Condorceta zawodów (zob. Rozdział 7.2), ale obowiązujący w tym turnieju system nie był niestety systemem condorcetowskim.

Takie paradoksalne wyniki spowodowały zmianę regulaminu. Podczas olimpiady w roku 2008 w Pekinie turniej zapaśniczy rozgrywano już systemem pucharowym, a olimpijskie złoto przyznawano, podobnie jak w boksie, niepokonanemu zawodnikowi. Zapaśnicy z jedną przegraną walką startowali w repasażach i mieli szansę walczyć o medal brązowy.

Omówiony przypadek bułgarskiego zapaśnika pokazuje, że dobrze, zdawałoby się, pomyślany regulamin zawodów może nie sprawdzić się w praktyce, gdyż dowolny matematycznie możliwy, nawet mało prawdopodobny, układ wydarzeń może kiedyś zajść w rzeczywistości. Podobny problem mógłby pojawić się w każdej dyscyplinie sportowej, w której wyróżnia się *różnie oceniane* formy zwycięstw. Przykładowo, w zawodach Pucharu Świata w **siatkówce** za każdą wygraną w stosunku 3:0 lub 3:1 przyznaje się obecnie 3 punkty, a za zwycięstwo 3:2 tylko 2 punkty. Łatwo wyobrazić sobie sytuację, w której drużyna A wygrałaby wszystkie swoje mecze, lecz w wielu przypadkach potrzebowałaby do tego aż pięciu setów i zgromadziłaby w konsekwencji mniej punktów niż drużyna B, która wprawdzie przegrałaby w bezpośrednim starciu z A,



lecz w pozostałych meczach zwyciężyłaby wysoko. W myśl regulaminu puchar zdobyłby wtedy zespół B z największym dorobkiem punktowym, choć pierwsze miejsce należałoby raczej przyznać niepokonanemu zespołowi A, condorcetowskiemu zwycięzcy zawodów.

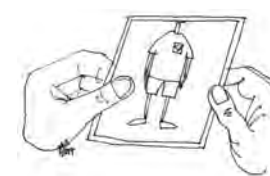
Opracowanie dobrego systemu rozgrywania turniejów sportowych wydaje się sprawą łatwą. Tak jednak nie jest, gdyż dość trudno czasem przewidzieć konsekwencje koincydencji kilku mało prawdopodobnych, choć możliwych zdarzeń. Wielokrotnie rozgrywane zawody w różnych dyscyplinach sportu stopniowo odsłaniają ułomności aktualnie stosowanych systemów punktacji, co sprzyja dalszej ewolucji regulaminów.

## Lekka atletyka i kwalifikacje na igrzyska olimpijskie

Miłośnicy królowej sportu rzadko mają kłopoty z ustaleniem, który zawodnik triumfuje w danej konkurencji. Wygrywa ten, kto skoczy najwyżej, rzuci najdalej lub jako pierwszy przybiegnie na metę biegu. Sytuacja komplikuje się nieco, gdy z grupy wielu dobrych **lekkoatletów** należy wybrać reprezentację kraju na igrzyska. W Stanach Zjednoczonych zwyczajowo sprawę załatwia się prosto: zawodnicy, którzy zajmą czołowe miejsca w mistrzostwach kraju, kwalifikują się do ekipy niezależnie od ich wcześniejszych osiągnięć. Na przykład, w biegu maratońskim *Olympic Trials* zorganizowanym 14 stycznia 2012 roku w Houston startować mógł każdy amerykański biegacz legitymujący się odpowiednio dobrym czasem. Do reprezentacji Stanów Zjednoczonych w maratonie na igrzyska olimpijskie w Londynie zakwalifikowali się trzej najlepsi uczestnicy.

Tymczasem w Polsce wybór składu reprezentacji kraju nie dokonuje się na stadionach, lecz w gabinetach Polskiego Związku Lekkiej Atletyki. Zasady kwalifikacji na drużynowe Mistrzostwa Europy 2011 stanowiły na przykład, że przy powoływaniu do reprezentacji Polski brane będą pod uwagę: a) wyniki osiągnięte przez zawodników w sezonie 2011, b) „narastające dyspozycje startowe”, c) „stan zdrowia”, a w niektórych konkurencjach także d) „cele strategiczne reprezentacji”. Taki sposób kwalifikacji przekazuje władzę w ręce działaczy sportowych, którzy mogą dowolnie interpretować powyższe reguły, ale z drugiej strony daje szansę utytułowanym sportowcom, którym nie powiódł się start w mistrzostwach kraju. Jak wyraził się kiedyś jeden z działaczy, trudno byłoby zaakceptować sytuację, że „na olimpiadę jada przypadkowi zwycięzcy zawodów o mistrzostwo Polski”.

Która metoda wyłaniania reprezentacji jest lepsza? Więcej sukcesów w lekkoatletyce z pewnością mają Amerykanie, ale też dzięki temu średni poziom zawodów lekkoatletycznych w Stanach Zjednoczonych jest wyższy i trudno tam „przypadkowo” zostać mistrzem kraju.



Tym razem aktualny mistrz kraju nie zmieścił się w kadrze

## Najlepszy poeta nigdy nie wygrywa<sup>9</sup>

**Slam poetycki** (od ang. *slam* – trzaskać), to inaczej konkurs recytujących poetów-performerów, którego zwycięzca otrzymuje wszystkie pieniądze za bilety wpłacone przez śledzącą występy publiczność; jest to więc dosłownie system *winner-takes-all*. Zasady rozgrywania tych zawodów stanowią oryginalną syntezę wielu metod głosowania przedstawionych w tej książce. Najpierw prowadzący slam wybiera spośród publiczności pięciu jurorów (co przypomina nieco losowanie sędziów w demokracji ateńskiej – Rozdział 1.1), którzy wartościują każdą recytację w skali od 0 do 10. Podobnie jak przy ocenianiu stylu w skokach narciarskich czy wartości technicznej programu w zawodach łyżwiarskich (Rozdział 11.1) dwie skrajne noty są odrzucane, a pozostałe sumuje się. Analogicznie natomiast jak w kolarskim wyścigu australijskim lub w wyborach prezydenta II Rzeczypospolitej (Rozdział 3.5) po każdej turze jeden z poetów odpada, a pozostali recytują kolejne wiersze, aż na placu boju pozostanie dwóch finalistów (niekiedy rozgrywane są też półfinały). O zwycięstwie w decydującej rozgrywce nie rozstrzyga już jednak jury, lecz cała zgromadzona publiczność, która, podobnie jak mieszkańcy starożytnej Sparty (Rozdział 1.2), po każdym z finałowych występów wyraża swoją ocenę brawami lub okrzykami.

Warto dodać, że pierwszy na świecie slam zorganizował w Chicago w roku 1984 robotnik budowlany a zarazem poeta **Marc Kelly Smith**. Moda na poetyckie slamy przysłała do naszego kraju w roku 2003, gdy pierwsza taka impreza odbyła się w Warszawie. Występujący w zawodach poeci nie tylko recytują swoje utwory, lecz także starają się za wszelką cenę przykuć uwagę publiczności – nie wolno im jednak używać w tym celu kostiumów, instrumentów czy rekwizytów, a występ może trwać co najwyżej trzy minuty. W tej „dyscyplinie” rozgrywane są też czasem zawody drużynowe na czele ze słynnym **National Poetry Slam** mającym rangę nieoficjalnych mistrzostw Stanów Zjednoczonych.



## 11.2. Konkurs Chopinowski i Triennale Grafiki

### 11.2.1. Międzynarodowy Konkurs Pianistyczny im. Fryderyka Chopina

Podobnie jak typowanie zwycięzców w sportach niewymiernych, tak i rozstrzyganie konkursów artystycznych stanowi nie lada problem. Zwykle wybór triumfatora pozostawia się decyzji fachowego jury. Musi ono jednak postępować zgodnie z regulaminem ustalonym przez organizatorów. W tym miejscu warto przypomnieć omówione dokładnie w Rozdziale 7 twierdzenia matematyczne wykluczające istnienie idealnego systemu wyłaniania zwycięzcy. Zapewne największym gwarantem poprawności decyzji jury jest dobór jego członków spośród najwybitniejszych fachowców z danej dziedziny. Przyjrzyjmy się zawiłościom takich regulaminów na przykładzie Międzynarodowego Konkursu Pianistycznego im. Fryderyka Chopina<sup>10</sup>. Ten cieszący się wielkim prestiżem konkurs po raz pierwszy odbył się w Warszawie w roku 1927. Od tej pory organizowany był co pięć lat, z wyłączeniem okresu II wojny światowej.



Mimo starannego doboru jurorów spośród światowej sławy pianistów i pedagogów dość często werdykt jury nie zgadza się z opinią publiczności<sup>11</sup>, a także krytyków muzycznych. Nie oznacza to oczywiście, że werdykt widowni jest trafniejszy, ale rozmijanie się decyzji jurorów z odczuciem publiczności wprowadza element dodatkowego, niekoniecznie przyjemnego napięcia. A przecież zwycięstwo w tak ważnym konkursie, chociaż samo przez się nie gwarantuje światowych sukcesów, to jednak takowe znacznie ułatwia.

W szesnastu zorganizowanych dotychczas edycjach konkursu cztery razy główne nagrody zdobywali pianiści polscy: Halina Czerny-Stefańska (1949), Adam Harasiewicz (1955), Krystian Zimerman (1975) oraz Rafał Blechacz (2005). W długiej historii konkursu zmieniała się nie tylko liczba uczestników i rozegranych etapów, lecz także sposób punktacji dokonania artystów i metoda wyłaniania zwycięzców. Także liczba przyznanych nagród nie zawsze była taka sama. Przykładowo, w roku 1949 wręczono dwie pierwsze nagrody *ex aequo*, a w latach 1990 i 1995 pierwszej nagrody jury nie przyznało wcale.

XVI Konkurs Chopinowski odbył się w październiku 2010 roku i brało w nim udział 81 pianistów. Konkurs tradycyjnie podzielony był na cztery etapy. W każdym z nich uczestnicy wybierali zestaw utworów Fryderyka

10) Analizę regulaminu tego konkursu z punktu widzenia teorii gier przeprowadziła w roku 2013 Honorata Sosnowska ze Szkoły Głównej Handlowej w Warszawie, która określiła obowiązujące reguły jako połączenie dwóch metod głosowania: punktowego i aprobowanego.

11) Taki pamiętny przypadek miał miejsce w roku 1980, gdy ulubieniec publiczności Ivo Pogorelić z Jugosławii nie został zakwalifikowany do finałowego etapu Konkursu.

9) Agata Kolodziej (red.), *Najlepszy poeta nigdy nie wygrywa. Historia slamu w Polsce 2003-2012*, Kraków 2013, to książka poświęcona slamowi w naszym kraju i dostępna w sieci pod adresem: <http://sc-ch.pl/najlepszy-poeta-nigdy-nie-wygrywa/>

Chopina z ustalonej listy, by w czwartym, finałowym etapie zagrać jeden z jego dwóch koncertów: e-moll lub f-moll<sup>12</sup>. W finale udział wzięło pięciu artystów z Rosji, dwoje z Francji, po jednym z Austrii i Bułgarii oraz jeden polski pianista, Paweł Wakarecy. Pierwszą nagrodę w konkursie zdobyła Julianna Awdiejewa (Rosja), a dwie równorzędne drugie nagrody zdobyli Lukas Geniušas (Rosja/Litwa) oraz Ingolf Wunder (Austria).

Według regulaminu do drugiego etapu konkursu miało awansować nie więcej niż 40 pianistów, do trzeciego – nie więcej niż 20, a do finału – nie więcej niż 10. W skład międzynarodowego jury oceniającego pianistów wchodziło 12 wybitnych muzyków, w tym 4 zwycięzców poprzednich edycji konkursu oraz 4 zdobywców dalszych nagród. Aby wykluczyć konflikt interesu, jurorzy musieli ujawnić, którzy z kandydatów są lub byli ich uczniami, i nie mogli brać udziału w ocenie ich występów. Nie jest jednak oczywiście prostą sprawą określenie granicy kontaktu z uczestnikiem konkursu, poza którą zasługuje on już na miano ucznia.

Podczas każdego etapu konkursu każdy juror przyznawał dwie oceny. Pierwsza ocena była prostą odpowiedzią „tak” lub „nie” na pytanie, czy kandydat powinien się znaleźć w następnym etapie konkursu. Odpowiedź „tak” należało postawić przy co najwyżej 40 nazwiskach w pierwszym etapie, 20 – w drugim i 10 – w trzecim. Liczba uzyskanych głosów „tak” istotnie wpływała na szanse awansu pianisty do kolejnego etapu, choć procedura wcale nie była prosta. Druga ocena przyznawana przez jurora była liczbowa, w skali od 1 (najniższa ocena) do 100 (ocena najwyższa).

Zauważmy tutaj, że położenie jurora w Konkursie Chopinowskim jest nieco bardziej komfortowe niż sędziego na zawodach gimnastycznych czy łyżwiarskich. O ile sędzia musi ujawnić swą notę natychmiast po wykonaniu danego ćwiczenia, przed startem kolejnego zawodnika, to oceniający grę pianistów może wystawić oceny po przesłuchaniu wszystkich uczestników danego etapu konkursu.

Z ocen liczbowych przyznanych artyście wyliczano średnią arytmetyczną. Aby wyeliminować wpływ skrajnych ocen, usuwano te, które różniły się od średniej więcej niż o 10, 8, 6 i 5 punktów w kolejnych etapach. Po usunięciu tych skrajnych ocen ponownie wyliczano pomocniczą średnią arytmetyczną. W przypadku skoków narciarskich w podobny sposób wyznacza się końcową ocenę, ale ocena konkursu pianistów jest bardziej skomplikowana. W tym punkcie ustalonej procedury skrajne oceny korygowane były tak, aby znalazły się na granicy dopuszczalnego przedziału wokół wyliczonej średniej pomocniczej. Po tym zabiegu raz jeszcze, już po raz ostatni, wyliczano średnią arytmetyczną, tym razem, jak na początku, używając wszystkich ocen.

Pełny zestaw ocen, jaki każdy uczestnik konkursu otrzymał w tym etapie od wszystkich jurorów był publicznie dostępny w witrynie internetowej



Julianna Awdiejewa, 1. nagroda w XVI edycji Konkursu Chopinowskiego w roku 2010

12) W historii konkursu tylko raz zwyciężył pianista, który wybrał koncert f-moll, co pozwala sądzić, że oba koncerty nie dają artyście takiej samej szansy na zademonstrowanie swych umiejętności.

13) Próbowano w ten sposób uzyskać kalibrację ocen wszystkich jurorów.

14) Zauważmy, że niniejsza książka też jest dziełem zbiorowym trzech autorów.

konkursu, co pozwoliło na ich analizę statystyczną. Z jednej strony podziw dla jury budzić może fakt, że niezależnie przyznawane oceny gry części pianistów niemal pokrywają się. Lecz z drugiej strony, w przypadku niektórych artystów, rozrzut przyznanych punktów jest znaczny, a znaleźć można też uczestnika konkursu, którego występ jeden juror ocenił na 95 punktów, a inny zaledwie na 20! Oczywiście te skrajne oceny nie weszły do obliczonej średniej oceny pianisty, ale podane liczby ilustrują, jak bardzo czasem różnią się odczucia członków jury i jak trudno jest wypracować wyważoną ocenę danego występu. Duży rozrzut ocen zmniejsza jednak szanse wykonawcy. W opinii krytyków regulamin konkursu faworyzuje więc raczej interpretacje poprawne, a nie oryginalne.

Jeżeli dobrze rozumiemy zawily język regulaminu, to głos „tak” liczył się tylko w przypadku, gdy dany juror przyznał pianiście co najmniej 75 punktów<sup>13</sup>. Teraz, po każdym etapie, sekretarz jury przedstawiał anonimowe wyniki kandydatów uporządkowane wedle liczby głosów „tak”, z dołączeniem średniego wyniku punktowego. Ten średni wynik punktowy kumulował się z etapu na etap, czyli był sumą średnich wyników punktowych z dotychczasowych etapów. Czytelnik zapewne sądzi, że teraz odpowiednie grono pianistów o największej liczbie głosów „tak” awansowało do kolejnego etapu. Otóż nie. Z wielkim zdumieniem czytamy, że o ostatecznej decyzji o promowaniu do kolejnego etapu decydowało każdorazowo jawne głosowanie jury!

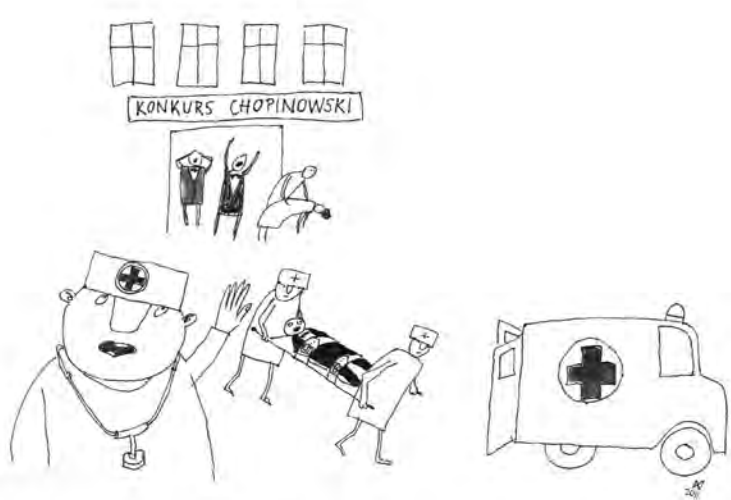
Po występach w finale sekretarz jury miał przedstawić, wciąż anonimowe, sumy liczbowych rezultatów poszczególnych pianistów i na ich podstawie jury miało rozdzielić nagrody. Regulamin znów nie narzucał automatyzmu, ale pozostawiał jurorom dużą swobodę.

Czytelnik, który z uwagą (lub bez) doczytał do tego miejsca opis regulaminu, zrozumie pewnie, dlaczego jurorzy konkursu zbuntowali się i wykorzystali przyznaną im swobodę decyzji. W finale konkursu jury ustaliło własny sposób liczbowej oceny pianistów, odmienny od narzuconego. Teraz każdy juror miał ocenić *cały* czteroetapowy występ finalisty w skali od 1 do 10. Tym razem, chyba aby zaakcentować swoją niezależność, jury przyjęło 1 jako ocenę najlepszą, a 10 – najgorszą. To średnia tych ocen decydowała ostatecznie o kolejności miejsc.

Przedstawiona tu historia ostatniego Konkursu Chopinowskiego jest bardzo pouczająca. Po przeczytaniu jego regulaminu chciałoby się zawołać: Autor! Autor! Pewnie było to ciało zbiorowe, którym kierowały najlepsze intencje. Jednakże dziwny w swojej złożoności efekt skłania do przypomnienia starego dowcipu: *Co to jest wielbłąd? To koń zaprojektowany przez komitet*<sup>14</sup>. System punktacji stosowany podczas Konkursu Fryderyka Chopina pokazuje dobitnie, jak trudno jest ustalić sensowne reguły wyboru zwycięzcy. Wielu ekspertów uważa ponadto, że znaczny

międzynarodowy prestiż Konkursu Chopinowskiego stopniowo obniża się w konsekwencji stosowanego regulaminu<sup>15</sup>, a wyniki konkursu byłyby mniej kontrowersyjne, gdyby niezależnie od możliwego uproszczenia systemu punktacji została wprowadzona reguła zabraniająca udziału w jury pedagogom, których uczniowie biorą udział w konkursie. Wydaje się bowiem, że samo powstrzymanie się jurorów od oceny własnych uczniów nie wystarczy.

*Naprawdę nie trzeba dziękować. Jak tylko usłyszałem, że w programie mają być jakieś wariacje, od razu podjęliśmy akcję prewencyjną.*



To prawdziwa przyjemność słuchać wspaniałej muzyki Chopina i oglądać przejętych konkursem pianistów, pozostając w błogiej nieświadomości co do sposobu wyłaniania zwycięzców. Obawiamy się, że tej nieświadomości naszych czytelników właśnie pozbawiliśmy.

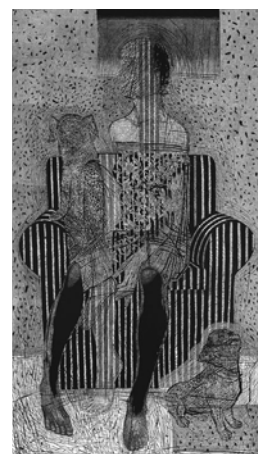
### Grand Prix Triennale Grafiki

Międzynarodowe Biennale Grafiki odbywało się w Krakowie od roku 1966. Od dwudziestu lat konkurs organizowany jest co trzy lata jako **Międzynarodowe Triennale Grafiki** (MTG). W pierwszym etapie konkursu MTG 2009 uczestniczyło 1643 artystów z 70 krajów, którzy nadesłali ponad 5500 prac. Do drugiego etapu zakwalifikowano 2201 prac 667 autorów, z których jury wybrało jedynie 300 do prezentacji podczas głównej wystawy. Wyniki konkursu ogłoszono 16 września 2009 roku, podczas uroczystości otwarcia Triennale, którą po raz ostatni poprowadził inicjator imprezy, profesor Witold Skulicz. W roku 2009 Grand Prix MTG zdobyła graficzka polska Joanna Piech-Kalarus.

Kolejne Triennale Grafiki odbyło się w roku 2012. W pierwszej fazie konkursu międzynarodowe jury oceniało prace nadesłane w formie elektronicznej. Autorzy prac zakwalifikowanych do drugiej fazy przestali

international print triennial  
in krakow

<sup>15</sup> Tymczasem rośnie prestiż innych konkursów pianistycznych w świecie, jak Międzynarodowego Konkursu Muzycznego im. Marii Canals w Barcelonie, w którym nie mogą startować uczniowie członków jury, a podstawy regulaminu da się streścić w jednej linijce: o awansie do dalszego etapu, a następnie o końcowym zwycięstwie, decyduje średnia ocen obliczona po odrzuceniu not skrajnych.



Linoryt *Wiem wszystko* Joanny Piech-Kalarus, laureatki Grand Prix MTG Kraków 2009

organizatorom oryginały swych prac, zgadzając się jednocześnie na ich wystawianie podczas Triennale lub towarzyszących mu wystaw. Członkowie jury selekcji obradowali w pięcioosobowym składzie w dniach 4–5 lipca 2012 roku i w tej fazie konkursu przejrzyli 2024 prace 1004 artystów. Do Wystawy Głównej zakwalifikowano prace 231 grafików. Jubileuszową, dwudziestą Wystawę MTG otwarto w krakowskiej Galerii Sztuki Współczesnej Bunkier Sztuki w dniu 14 września 2012 roku. Nagrodę Grand Prix d'Honneur zdobyła Izabella Gustowska za pracę *Hybrydy 1, 2, 3*, a nagrodę Grand Prix – Joanna Janowska-Augustyn za dzieło *Kolej rzeczy I*.

Do udziału w międzynarodowym jury nagród organizatorzy Triennale zapraszają uznanych artystów z całego świata, a także fachowych krytyków sztuki, kuratorów wystaw oraz historyków sztuki. Jak to zróżnicowane grono ustala, który artysta zdobędzie Grand Prix?

Początkowo dominowało niespieszne ucieranie poglądów w wyniku prowadzonych między jurorami dyskusji i sporów. Podczas ostatnich edycji Triennale międzynarodowe jury działa na nieco bardziej sformalizowanych zasadach. Po dwudniowym przeglądzie wszystkich prac nadesłanych na II fazę konkursu i kwalifikacji części dzieł do głównej wystawy, w ciągu jednego dnia pracy wybiera ono zdobywców głównych nagród. Najpierw każdy juror typuje grafiki do finałowej grupy około 30–40 prac, wśród których poszukiwać się będzie zwycięzcy. Następnie członkowie jury otrzymują po 5 kartek do głosowania i oddzielnie spacerują po salach, gdzie leżą uprzednio wyłonione prace. Każdy juror zostawia swój głos na pięciu grafikach, które typuje do nagród konkursu.

*Mówiłem panu, mam niepodważalny dowód na to, że Picasso nie był abstrakcjonistą.*



Po obliczeniu wyników tego wstępnego głosowania rozpoczyna się dłuższa debata, w której jurorzy wymieniają opinie oraz formułują argumenty za poszczególnymi dziełami lub przeciw nim. Dopiero po tej dyskusji i odrzuceniu prac, które otrzymały małą liczbę kartek, przeprowadza się kolejne głosowanie nad przyznaniem Grand Prix i innych nagród Triennale. Biorąc pod uwagę jego wyniki, przygotowuje się końcowe orzeczenie jury, pod którym podpisują się wszyscy jego członkowie. W oczywisty sposób na ostateczny wynik mają zatem szczególnie wpływ jurorzy dysponujący wyjątkowym autorytetem oraz odpowiednią siłą perswazji.

Wprawdzie łyżwiarstwo figurowe jest sportem niewymiernym, ale można opracować listę błędów przy wykonywaniu ewolucji z odpowiadającą im skalą odejmowanych punktów. Ocena pracy artysty jest sprawą znacznie bardziej subtelną i bezpośrednio przeliczanie klasy dzieła na punkty nie wydaje się celowe. Jak widzieliśmy próby zbyt sformalizowanej oceny punktowej uczestników Konkursu Chopinowskiego natrafiają na opór jurorów. Niemniej podstawowy problem teorii społecznego wyboru – jak przełożyć zróżnicowane opinie wielu jurorów na jedną wspólną decyzję jury – pozostaje równie aktualny w konkursach sztuki, jak i w zawodach sportowych.

## 11.3. Nagrody Premiera i granty europejskie

### 11.3.1. Który uczony polski otrzyma doroczną Nagrodę Premiera?

W roku 1994 na wniosek Polskiej Akademii Nauk ustanowiono doroczne **Nagrody Premiera** dla wyróżniających się polskich naukowców. Od tego czasu przeżyliśmy już liczne rządy i wielu premierów, ale każdy z nich kontynuuje rozpoczętą tradycję. W roku 2011 nagrody przyznano w pięciu kategoriach: za wyróżnione prace doktorskie (kandydatom do 30. roku życia), wyróżnione rozprawy habilitacyjne (kandydatom do 38. roku życia), wybitne osiągnięcia naukowe lub artystyczne, wybitne osiągnięcia naukowo-techniczne oraz za wybitny dorobek naukowy.

Dnia 28 marca 2011 roku w Centrum Nauki Kopernik w Warszawie premier Donald Tusk wręczył nagrody 46 naukowcom wyróżnionym w roku poprzednim. Przyznano pięć nagród za wybitny dorobek naukowy lub artystyczny oraz trzy za wybitne osiągnięcia naukowe. Ponadto wręczono 9 nagród za wyróżniające się habilitacje, 28 – za wyróżniające się prace doktorskie oraz 3 nagrody za wybitne osiągnięcia naukowo-techniczne. Nagroda za wyróżniony doktorat wynosiła 25 tysięcy złotych, podczas gdy nagrody w innych kategoriach były jeszcze wyższe, co dodatkowo podnosi prestiż Nagrody Premiera w środowisku polskich naukowców.

Ale jak ocenić pracę naukowców i porównać na przykład doktoraty z biologii molekularnej i filozofii lub habilitacje z medycyny i socjologii? Czy da się obiektywnie ocenić dorobek wybitnego astronoma, aby stwierdzić, że zasługuje on bardziej na nagrodę niż znany historyk? Odpowiedź jest prosta: nie da się! Jednak kandydatów do Nagrody Premiera w jakiś sposób wybrać trzeba...

Nominacje przygotowuje powoływany przez premiera zespół naukowców, który w latach 2010–2012 liczył 26 osób reprezentujących różne dziedziny nauki i różne ośrodki akademickie. Zespół podzielono na cztery podzespoły grupujące przedstawicieli nauk humanistycznych i artystycznych, nauk ścisłych, nauk przyrodniczo-medycznych oraz nauk technicznych, które niezależnie oceniały wnioski z danej dziedziny nadsyłane przez poszczególne uczelnie lub instytuty naukowe.

Podczas pierwszego spotkania w Warszawie każdy podzespół przejrzał wszystkie przekazane mu wnioski i dla każdego z nich wyznaczył recenzenta spośród swych członków. Przewodniczący zespołu zaproponował wstępny podział całkowitej liczby nagród (na przykład 25 nagród za doktoraty) na poszczególne podzespoły, biorąc pod uwagę liczbę kandydatów zakwalifikowanych w każdej kategorii do danej grupy nauk.



Uroczystość wręczenia Nagród Premiera w Centrum Nauki Kopernik, marzec 2011

Każdy oceniający działał według ściśle sformułowanych reguł, oceniając doktoraty w skali od 0 do 12 punktów, według kryteriów ustalonych w danym podzespołe. Wnioski zawierały wiele załączników ułatwiających pracę recenzenta, jak opinie Komitetów Polskiej Akademii Nauk z danej dziedziny wiedzy. Pomimo tych narzędzi podzespoły nadal miały problem, jak porównać doktorat chemika i matematyka, skoro każdy recenzent jest kompetentny tylko w jednej dziedzinie nauki.

Kluczowymi elementami działania zespołu były wzajemne zaufanie jego członków i próba wspólnej kalibracji używanej skali punktowej. Drugie spotkanie całego zespołu w Warszawie rozpoczęło się równoległymi obradami w czterech podzespołach. Biorąc pod uwagę wszystkie recenzje oraz oceny punktowe, każdy podzespół wspólnie wypracował listy rankingowe swych kandydatów w każdej kategorii. W początkowej fazie obrad, gdy matematycy stanowczo przekonywali o wybitnych doktoratach z matematyki, chemicy zachwalali chemików, a fizycy – fizyków, wydawało się, że nie ma dużych szans na osiągnięcie kompromisu. Ale po pewnym czasie trwania sporu, gdy wszystkie listy punktowe w każdej kategorii były już znane, najczęściej do porozumienia dochodziło na drodze negocjacji, a tylko sporadycznie w podzespołach przeprowadzano formalne głosowanie.

Następnie na posiedzeniu plenarnym przewodniczący każdego podzespołu przedstawiał kolejne kandydatury w danej kategorii nagród. Propozycje mieszczące się w ustalonej puli nagród, która przypadała na dany podzespół, zostawały dość szybko przegłosowane i rozpoczynała się najtrudniejsza część dyskusji na temat pozostałych nagród. W tym momencie trzeba było znów porównać nieporównywalne i rozstrzygnąć, czy wnioskować o nagrodę za odkrycie archeologa, czy też za udowodnienie twierdzenia matematycznego. W toku trudnych negocjacji dochodziło do wypracowania kompromisu.



Czy stosowana metoda przyznawania nagród jest w pełni sprawiedliwa? Z pewnością nie, gdyż jeśli w ciągu jednego roku pięciu młodych chemików obroni w Polsce doktoraty światowej klasy, to i tak nie wszyscy z nich otrzymają Nagrodę Premiera. Ale też procedura selekcyjna sprawia, że nagród nie otrzymują kandydaci słabi lub średni. Fakt, że wielu bardzo silnych kandydatów nie zostanie nagrodzonych, jest łatwiejszy do przyjęcia dla całego zespołu oraz dla środowiska naukowego.

Spoglądając na listę laureatów Nagród Premiera oraz na wyniki innych konkursów naukowych w Polsce, łatwo dostrzec, że przyjęte procedury mają kilka trudnych do uniknięcia wad, powiązanych z fragmentacją różnych dziedzin nauki. Poszczególne poddziedziny tworzą grupy nacisku, których oddziaływanie zależy raczej od liczebności ich przedstawicieli niż od jakości danej specjalności. W dodatku środowiska naukowe w Polsce nie są wielkie i wszyscy liczący się reprezentanci danej dziedziny zwykle dobrze się znają. Ze względu na dotychczasowe znajomości jedni się popierają, a inni – zwalczają. W nauce panują też zmienne mody, podobnie jak w innych dziedzinach życia. Uprawianie problematyki modnej w danym czasie problematyki niewątpliwie ułatwia sukces.

W wynikach niektórych konkursów widać też wpływ politycznej poprawności, wymagającej na przykład przyznania pewnej liczby wyróżnień naukowcom z mniejszych ośrodków. Przy konieczności wyboru zaledwie jednego lub dwóch laureatów w danej kategorii ze znacznej grupy kandydatów pewne znaczenie ma także czynnik losowy. Ponadto nawet pełne najlepszej woli gremia oceniające często ulegają jurorom o szczególnie silnej osobowości. Konkurs jest więc nie tylko polem rywalizacji między kandydatami do nagrody, lecz także między osobami oceniającymi, zwłaszcza w fazie bezpośrednich dyskusji.



### Europejskie Granty ERC ideas

Aby stymulować prowadzenie badań naukowych na światowym poziomie, Unia Europejska finansuje prestiżowe **granty badawcze**, o które na zasadzie konkurencji mogą ubiegać się naukowcy z całej Europy. Granty przyznawane są obecnie corocznie w trzech głównych kategoriach: *Starting Grants* – dla młodych badaczy, do 7 lat po doktoracie, *Consolidator Grants* – dla naukowców, którzy tytuł doktora uzyskali od 7 do 12 lat wcześniej oraz *Advanced Grants* – stanowiące kategorię otwartą dla najlepszych uczonych całego kontynentu. Wnioski przyjmuje i granty przyznaje **European Research Council (ERC)** z siedzibą w Brukseli, w skład której wchodzi znani uczeni reprezentujący różne dziedziny nauki i kraje Europy. W roku 2011 złożono 4080 wniosków o granty ERC *Starting*, a przyznano ich 480. Średnia wartość realizowanego grantu wynosiła 1,4 miliona euro, co miało pozwolić laureatom konkursu na stworzenie własnej grupy badawczej i sfinansowanie jej działalności przez okres 3–5 lat.

Od momentu rozpoczęcia programu w roku 2007 do końca roku 2013 przyznano w tych kategoriach niemal 3400 grantów ERC. Na razie lista polskich laureatów tych konkursów jest bardzo krótka<sup>16</sup>: w kategorii *Advanced* wyróżniono projekty Tomasza Dietla (fizyka) i Andrzeja Udałskiego (astronomia) z Warszawy oraz Ryszarda Horodeckiego (fizyka,

<sup>16</sup> Jako punkt odniesienia dodajmy, że latach 2007–2011 uniwersytety w Cambridge i Oxfordzie otrzymały odpowiednio 76 oraz 72 granty ERC, natomiast uniwersytety w Leuven (Belgia), Lejdzie i Amsterdamie (Holandia) oraz Helsinkach (Finlandia) ponad 20 grantów każdy, a więc znacznie więcej niż zdobyli w tym okresie łącznie wszyscy naukowcy z Polski.

Gdańsk). Takie prestiżowe granty otrzymali też uczeni polscy pracujący za granicą: Maciej Lewenstein (dwukrotnie! – fizyka, Barcelona), Andrzej Buras (fizyka, Monachium), Piotr Smolarkiewicz (geofizyka, Reading k. Londynu), Agata Smoktunowicz (matematyka, Edynburg), Andrzej Stankiewicz (chemia, Delft), oraz Jerzy Paszkowski (biologia, Genewa). W kategorii *Starting* granty otrzymali: Mikołaj Bojańczyk, Stefan Dziembowski i Piotr Sankowski (informatyka), Janusz Bujnicki (genetyka), Andrzej Dziembowski i Marcin Nowotny (biologia molekularna), Piotr Garstecki (chemia fizyczna), Natalia Letki-Garner (socjologia), Justyna Olko-Bajer (historia), Piotr Sułkowski (fizyka) – wszyscy z Warszawy – oraz Maciej Konacki (astronomia, Toruń/Poznań).

Jak wyglądał proces selekcji wniosków i wybór tych najlepszych, zakwalifikowanych do finansowania? W określonym terminie każdy wnioskodawca składał swój projekt w formie elektronicznej przez witrynę internetową ERC. Złożone wnioski zostawały przyporządkowane wstępnie do jednej z sekcji tematycznych, odpowiadającej wcześniej powołanym panelom kilkunastu europejskich ekspertów z danej dziedziny. Następnie szef każdego panelu, którym przeważnie był znany naukowiec, oraz pomagający mu urzędnik ERC (najczęściej z doktoratem z danej dziedziny wiedzy), rozdzielał poszczególne projekty między członków panelu. Każdy wniosek o *Advanced Grant* oceniany był przez czterech ekspertów, przy czym dokładano starań, aby dwóch z nich specjalizowało się w tej samej dziedzinie, co wnioskodawca. Dwóch innych ekspertów miało za zadanie ocenić, czy projekt badawczy jest sformułowany na tyle jasno, że o celowości projektu mogą zaświadczyć także przedstawiciele innych specjalności.

Na każdego eksperta wypadało w tej fazie średnio około czterdzieści wniosków do oceny, więc oceniający, miał dobrą skalę odniesienia. Członek panelu otrzymywał elektronicznie pełny wniosek kandydata i pracując w domu, miał ocenić w skali od zera do czterech punktów dwa aspekty wniosku:

- a) dotychczasowe osiągnięcia naukowe kandydata,
- b) wartość merytoryczną projektu.

Przykładowo, recenzent miał odpowiedzieć na pytanie, czy dotychczasowy dorobek kandydata z ostatnich dziesięciu lat pozwolił mu uzyskać reputację lidera w swej dziedzinie oraz czy proponowany projekt jest dostatecznie ambitny i wychodzi poza aktualny stan badań w danej dyscyplinie. Suma średnich ocen udzielanych w punktach a) i b) przez wszystkich oceniających pozwoliła ERC na przygotowanie wstępnej listy rankingowej wszystkich wniosków dla danego panelu. Pozornie obie oceny są zupełnie niezależne, ale praktyka pokazuje, że średnie obu ocen są silnie skorelowane. Wysokie oceny wniosku otrzymywali zazwyczaj



kandydaci o znacznym dorobku, wielu wcześniej zrealizowanych grantach i ugruntowanej pozycji w środowisku.

Każdemu wnioskowi przydzielano jednego uczestnika panelu (ang. *lead reviewer*), który przedstawiał projekt badawczy podczas pierwszego posiedzenia panelu w Brukseli. Szczegółowe dyskusje nad każdym projektem przeciągały się, a efektem prac trzydniowego posiedzenia panelu było sporządzenie nowej listy ocen wszystkich wniosków, która wcale nie różniła się tak bardzo od wstępnej listy rankingowej. Mniej więcej trzecia część najlepiej ocenianych wniosków otrzymała ocenę A i została zakwalifikowana do drugiego etapu konkursu.

W drugiej fazie konkursu każdy wniosek kategorii A bardziej szczegółowo oceniali wybrani członkowie panelu, którzy nie opiniowali go w pierwszym etapie, a także wybrani przez panel zewnętrzni eksperci, niekoniecznie pracujący w Europie. Średnia wszystkich ocen stanowiła podstawę do stworzenia nowej listy rankingowej. Potem członkowie panelu spotykali się po raz drugi na trzydniowym posiedzeniu w Brukseli. Tym razem owocem ich pracy była końcowa lista wniosków, pod którą podpisywali się wszyscy członkowie panelu. Szczegółowe decyzje wypracowywano w ciągu wielogodzinnych dyskusji, w trakcie których rzadko dochodziło do formalnego głosowania. Na około 100–150 wniosków rozpatrywanych w danej grupie, ERC zdecydowała się finansować kilkanaście według listy ustalonej przez panel.

W pierwszej fazie konkursu nie brano pod uwagę części wniosku dotyczącej finansowania projektu. Dopiero w drugim etapie konkursu omawiano szczegóły finansowe i tylko niekiedy sugerowano zmniejszenie budżetu grantu. W podobny sposób przebiegała też ocena wniosków o granty w kategorii *Starting* dla młodszych badaczy, z jedną istotną różnicą: podczas II etapu konkursu autorzy najlepszych wniosków zapraszani byli do Brukseli, aby osobiście przedstawić swój projekt członkom panelu ekspertów.

Czy opisany system oceny wniosków jest obiektywny i sprawiedliwy? W zasadzie tak, choć mimo znacznego nakładu pracy ekspertów pewną rolę w systemie odgrywa czynnik losowy. Oczywiście wszyscy eksperci nie mogą recenzować wszystkich wniosków, więc kluczowym punktem procedury jest przydzielanie konkretnych wniosków konkretnym członkom panelu, którzy reprezentują różne dziedziny danej gałęzi nauki. W tej sytuacji większe szanse na końcowy sukces mają uczeni, których wyniki są powszechnie znane także poza ich wąską specjalnością badawczą. Podobnym problemem jest znalezienie w drugim etapie konkursu kompetentnych recenzentów zewnętrznych, którzy zgodziliby się zrecenzować dany wniosek w określonym terminie, oraz sprawienie, że ich oceny punktowe będą spójnie wykalibrowane. Specyfiką konkursu jest

konieczność wybrania do finansowania jedynie około 10% wniosków. W takich warunkach nie sposób uniknąć znacznej roli przypadku.

Zamierzeniem ERC jest finansowanie badań najlepszych uczonych naszego kontynentu. Stosowany system selekcji jest skuteczny o tyle, że wnioski słabszych kandydatów nie mają żadnych szans na sukces. Laureatami konkursów ERC zostają znani naukowcy z wielkim dorobkiem i projektami nowatorskich badań. Z drugiej strony, konkurencja w obu kategoriach jest bardzo ostra i wielu świetnych badaczy ze wspaniałymi osiągnięciami przepada w rywalizacji. Zarówno w Polsce, jak i w całej „Nowej Europie” nie ma obecnie wielu badaczy zdolnych do konkurowania z najlepszymi w Unii Europejskiej, więc na istotnie lepsze wyniki w konkursach ERC musimy jeszcze poczekać. W dodatku występuje tu pewne ujemne sprzężenie zwrotne: niedoinwestowana nauka Europy Środkowo-Wschodniej ma niewielu uczonych na najwyższym światowym poziomie i dlatego jest także niedostatecznie reprezentowana wśród międzynarodowych grup recenzentów oraz członków paneli ekspertów. Lamentując nad słabymi wynikami polskich naukowców, warto pamiętać, że wśród zwycięzców konkursów ERC jest także bardzo mało uczonych z innych peryferyjnych krajów Unii Europejskiej, takich jak Portugalia czy Grecja.



Porsche Cayman i Lotus Evora

## 11.4. Rankingi samochodów, uniwersytetów, miast i gier planszowych

### Najlepszy sportowy samochód roku

Amerykańskie czasopismo poświęcone motoryzacji „Car and Driver” corocznie przeprowadza testy nowych pojazdów różnych klas. W roku 2010 najlepsze **samochody sportowe** testowano podczas długiego rajdu przez wymagające przełęcze Gór Skalistych oraz na torze samochodowym w okolicy Los Angeles. Ekspertsi oceniali każdy samochód pod względem pięciu różnych kryteriów. Biorąc pod uwagę średnią z tych ocen, organizatorzy ogłosili następujące wyniki rankingu, w którym każdy samochód mógł uzyskać co najwyżej 250 punktów:

1. **Porsche Cayman** 193 pkt,
2. Chevrolet Corvette 186 pkt,
3. Lotus Evora 182 pkt.

W komentarzu redakcyjnym pisma, bardzo wpływowego na amerykańskim rynku motoryzacyjnym, podsumowano przebieg przeprowadzonych testów i określono samochód Porsche Cayman terminem *clear winner*, czyli „wyraźny zwycięzca”, całej konkurencji.

Jak podkreślają uważni obserwatorzy, metodologia tworzenia rankingu samochodów sportowych była dokładnie taka sama, jak równolegle prowadzonego rankingu popularnych samochodów klasy kompaktowej. W tej kategorii waga czynnika: „wygląd pojazdu” nie wydaje się bardzo istotna, ale może być decydująca dla zamożnych klientów rozważających zakup auta sportowego. Jak wskazał Malcolm Gladwell w artykule opublikowanym w roku 2011 w piśmie „New Yorker”, gdyby uwzględnić te same oceny wszystkich ekspertów, lecz zwiększyć wagę czynnika „wygląd pojazdu” oraz „przyjemność z jazdy” to wynik testu wyglądałby następująco:

1. **Lotus Evora** 205 pkt,
2. Porsche Cayman 198 pkt,
3. Chevrolet Corvette 192 pkt.

W algorytmie używanym przez redakcję „Car and Driver” cena pojazdu wpływa na końcową ocenę rankingu z wagą jedynie 8%. Może dla najbogatszych fanów motoryzacji przy zakupie samochodu sportowego cena istotnie nie odgrywa wielkiej roli, ale dla wielu innych potencjalnych klientów różnica 13 tysięcy dolarów, bo o tyle tańszy jest Corvette od Caymana, ma spore znaczenie. Jeśliby zmodyfikować względne wagi poszczególnych ocen, a 33% punktów przypisać cenie samochodu, to na czele rankingu pojawiłby się Corvette,



1. **Chevrolet Corvette** 205 pkt,
2. Lotus Evora 195 pkt,
3. Porsche Cayman 195 pkt.

Ostatecznie który więc samochód sportowy był „najlepszy” w roku 2010? Pomimo bardzo zdecydowanych sformułowań redaktorów pisma „Car and Driver” wydaje się, że przeprowadzony ranking nie pozwala udzielić jasnej odpowiedzi na to pytanie. Przeprowadzone testy mogą budzić zaufanie, a wyniki rankingu umożliwiają odróżnienie samochodów bardzo dobrych od tych wyraźnie gorszych, ale nie pozwalają na wyłonienie wyraźnego zwycięzcy, tym bardziej że znaczenie sformułowania „najlepszy samochód” nie jest jednoznaczne. Ponieważ grupa kilku czołowych pojazdów jest bardzo wyrównana, pierwsze miejsce w klasyfikacji może przypaść dowolnemu samochodowi z pierwszej trójki, w zależności od ustawienia wag, z jakimi bierze się pod uwagę różne aspekty ocenianego pojazdu. Krótko mówiąc, taki ranking nie jest w pełni obiektywny, gdyż przy próbie jednoczesnego zastosowania kilku niezależnych kryteriów ostateczny wynik silnie zależy od decyzji osób opracowujących schemat punktacji.

### Który uniwersytet jest najlepszy?

Podobne problemy dotyczą wielu rozmaitych rankingów, w których końcowa klasyfikacja ma odzwierciedlać wiele kryteriów. W ostatnich latach sporo uwagi przywiązuje się przykładowo do **rankingów uniwersytetów**, zarówno w skali Polski, Europy, jak i świata. W Polsce popularny jest **ranking szkół wyższych** publikowany przez „Perspektywy” i „Rzeczpospolitą”, który ocenia siłę naukową uczelni, jej prestiż mierzony preferencjami pracodawców i opiniami samej kadry naukowej, warunki studiowania, umiędzynarodowienie studiów oraz innowacyjność. W latach 2010 i 2011 pierwsze miejsce w tym rankingu zajęł **Uniwersytet Warszawski** przed **Uniwersytetem Jagiellońskim**. W następnych dwóch latach kolejność była odwrotna, jednak różnice między dwoma czołowymi polskimi uczelniami pozostawały minimalne: w ostatniej edycji zaledwie 0,08 punktu na 100 możliwych. Kolejne miejsca w rankingu w roku 2013 zajęły Uniwersytet im. Adama Mickiewicza w Poznaniu, Politechnika Warszawska oraz Akademia Górniczo-Hutnicza w Krakowie.

Wśród podobnych przedsięwzięć oceniających uniwersytety w skali całego świata najlepiej znany jest ranking **ARWU** (*Academic Ranking of World Universities*) przygotowywany corocznie, począwszy od roku 2003, przez Uniwersytet Jiao Tong w Szanghaju i dlatego potocznie nazywany **rankiem szanghajskim**. Jego pierwotnym celem było dostarczanie wiarygodnych informacji chińskim studentom zamierzającym podjąć studia za granicą. Ranking uwzględni kilka dość dyskusyjnie dobranych



kryteriów: liczbę laureatów Nagrody Nobla z nauk ścisłych oraz zdobywców Medalu Fieldsa z matematyki wśród wychowanków i profesorów uniwersytetu, liczbę autorów licznie cytowanych prac naukowych, liczbę artykułów opublikowanych w prestiżowych czasopismach „Nature” i „Science” w ciągu ostatnich pięciu lat, całkowitą liczbę wszystkich prac opublikowanych przez pracowników uniwersytetu w minionym roku i indeksowanych w bazach firmy Thomson ISI (czyli na tak zwanej liście filadelfijskiej) oraz iloraz średniego wskaźnika uniwersytetu w powyższych kategoriach do liczby jego pracowników. W roku 2012 w rankingu zwyciężył **Uniwersytet Harvarda** przed Uniwersytetem Stanforda i Massachusetts Institute of Technology (MIT), a w pierwszej dziesiątce było aż osiem uniwersytetów amerykańskich. Najwyżej sklasyfikowanymi uniwersytetami w Europie były Cambridge, a za nim Oxford, które uzupełniały pierwszą dziesiątkę listy. Wśród pierwszych 400 uniwersytetów świata znalazły się jedynie dwa uniwersytety polskie (UW i UJ) sklasyfikowane *ex aequo* na miejscach 301–400.

W nieco odmiennym rankingu **Times Higher Education World University Ranking** oceniano między innymi nauczanie, badania naukowe, stosunek liczby cytowań do liczby artykułów publikowanych przez pracowników uniwersytetu oraz międzynarodową reputację uczelni. Wedle danych za rok 2013 najwyżej ocenioną uczelnią na świecie był **California Institute of Technology** (Caltech) z wynikiem 94,9 punktów, przed Harvardem i Oxfordem, które uzyskały po 93,9 punktu i minimalnie wyprzedziły Stanford. Z uniwersytetów polskich uwzględniono tylko jeden (UW), który zajął miejsce od 301 do 350. W hiszpańskim rankingu **Webometrics**, bazującym na dokumentach zamieszczonych w Internecie, w edycji 2013 sklasyfikowano ponad 20 000 szkół wyższych<sup>17</sup>. Na czele znajdują się Harvard, Stanford i MIT, a całą pierwszą dziesiątkę zajmują uniwersytety ze Stanów Zjednoczonych. W tym rankingu Uniwersytet Warszawski znalazł się na 234 miejscu, a Jagielloński – na miejscu 311.

Jakie wnioski można wyciągnąć z tych danych? We wszystkich rankingach uniwersytetów kilka zupełnie różnych kryteriów charakteryzujących daną uczelnię próbowano opisać za pomocą jednego wskaźnika. Podobnie jak w przypadku klasyfikacji samochodów sportowych, ostateczne wyniki rankingu będą istotnie zależały od wyboru poszczególnych kryteriów oraz wag, z jakimi dane oceny wchodziły do końcowego wyniku<sup>18</sup>. Dlatego też na podstawie rezultatu rankingów nie da się stwierdzić, czy Harvard jest „lepszy” od Stanforda ani czy w Europie Oxford wyprzedza Cambridge. Żaden ranking w Polsce w przekonujący sposób nie wykaże, czy „lepszy” jest Uniwersytet Warszawski czy Jagielloński, gdyż różnice pomiędzy nimi są mniejsze niż dokładność, z jaką przyznaje się punkty. Inaczej mówiąc, nietrudno jest skonstruować pozornie „obiektywny” ranking uczelni

17) Zob. <http://www.webometrics.info>.

18) Z geometrycznego punktu widzenia, jeżeli rzutujemy punkty danego wielowymiarowego zbioru na jedną oś liczbową, to uporządkowanie ich rzutów na osi zależy od wyboru kierunku rzutowania.

polskich, w którym na czele tabeli pojawi się UW lub UJ, wedle życzenia zamawiającego. Z drugiej strony w różnie tworzonych na świecie rankingach regularnie dominują te same uczelnie amerykańskie, w Europie zaś najwyższe oceniane są uczelnie brytyjskie. W każdym rozsądnym<sup>19</sup> rankingu Oxford będzie stał znacznie wyżej niż UW, a Cambridge<sup>20</sup> – wyżej niż UJ. Nie przywiązując nadmiernego znaczenia do wyników jednego czy drugiego rankingu, nie powinno się też całkowicie lekceważyć faktu, że na corocznie aktualizowanych listach najlepszych uczelni świata polskie szkoły wyższe zajmują tak odległe miejsca.

## W którym mieście żyje się najlepiej?

Amerykańska firma konsultingowa Mercer ocenia **jakość życia** w najważniejszych **miastach** całego świata. Pod uwagę bierze się między innymi czynniki ekonomiczne, polityczne, klimatyczne, bezpieczeństwo mieszkańców, system szkolnictwa, opiekę medyczną, transport publiczny, życie kulturalne, zaopatrzenie i łatwość wynajęcia mieszkania. W rankingu Mercer za rok 2012 pod względem jakości życia zwyciężył **Wiedeń** przed Zurychem, Auckland, Monachium i Vancouver. Można dyskutować na temat tego werdyktu, bo gdyby nieznacznie zmodyfikować wagi, z jakimi różne czynniki wchodzi do końcowej oceny, kolejność tej piątki mogłaby się zmienić, ale łatwo się zgodzić, że czołówka rankingu wyprzedziła Pragę (69. miejsce), Budapeszt (74.), Wilno (79.) oraz **Warszawę** (84.). Z drugiej strony, w tych bliższych nam geograficznie miastach z pewnością żyje się łatwiej niż w Tbilisi (Gruzja – 213. miejsce), Chartumie (Sudan, 218.) lub wycieńczonym wojną **Bagdadzie**<sup>21</sup> (Irak), który zajął ostatnią, 221. pozycję rankingu. Różnica jednego miejsca w klasyfikacji nie wydaje się znacząca, ale ranking jest w stanie dobrze rozróżnić metropolie, w których żyje się łatwo, trudniej i wreszcie – bardzo trudno.

## W jaką grę planszową warto zagrać?

W przypadku stosowania tylko jednego kryterium wybór najlepszego elementu ze zbioru kandydatów staje się prostszy: wygrywa kandydat z największą liczbą punktów lub najwyższą średnią ocen. Sytuacja nieco się komplikuje, gdy na kolejnych kandydatów oddano inną liczbę głosów: czy wyżej cenić jedną ocenę na 10 punktów czy sto ocen ze średnią 9,0?

Podobnie jest z wieloma rankingami internetowymi różnych produktów, w których liczba ocen każdego modelu jest inna. Na stronie internetowej **BoardGameGeek (BGG)** można oceniać najlepsze **gry planszowe** sprzedawane na całym świecie. Niektóre gry zostały już ocenione przez ponad 30 tysięcy graczy, a każdy z nich stawiał ocenę w skali od 1 do 10. Oczywiście łatwo jest wyliczyć średnią ocenę i posegregować gry według takiej średniej. Lecz jeśli na rynek wchodzi nowa



19) Jako przykład mniej rozsądnego rankingu można sobie łatwo wyobrazić użycie kryterium „liczba profesorów polonistyki na uniwersytecie”, które wchodzi do oceny końcowej z wagą 99%...

20) W rankingach uczelni brytyjskich Oxford i Cambridge wyraźnie wyprzedzają pozostałych konkurentów.

21) Smuci obecny los Bagdadu, miasta *Baśni z tysiąca i jednej nocy*, który w wiekach IX–XI był znanym ośrodkiem kultury oraz światową stolicą matematyki i astronomii. Gdyby Mercer organizował swój ranking tysiąc lat wcześniej, być może wygrałby go Bagdad przed Kordobą...



Gry planszowe *Dominion: Intrigue* oraz *War of the Ring*

Wybrane gry planszowe oraz ich ranking BGG: stan z grudnia 2011

22) **Thomas Bayes** (1701–1761) – angielski teolog i matematyk, autor twierdzenia o prawdopodobieństwie warunkowym nazywanego dziś twierdzeniem Bayesa.

23) Ta liczba jest jedynym parametrem modelu i zależy od oczekiwanej liczby głosujących.

24) Podejście Bayesa wykorzystane jest na przykład w znanym rankingu filmów na portalu [www.filmweb.pl](http://www.filmweb.pl).

gra i początkowo zostaje ona bardzo wysoko oceniona przez zaledwie kilku graczy, to jak porównać taką grę z inną, ocenioną nieco niżej przez tysiące graczy? Czy wyżej ocenić grę *Dominion: Intrigue* z roku 2009, dla której średnia z 9335 ocen wynosi 8,03, czy też nowszą grę *War of the Ring*, rocznik 2010, ocenioną przez 738 graczy średnio na 8,69 punktów?

Konkretnej odpowiedzi na takie pytanie dostarcza podejście Bayesa<sup>22</sup> do zagadnień statystycznych, w którym zakłada się, że prawdopodobieństwo można interpretować jako stopień naszej wiedzy o określonym zagadnieniu. Dlatego nie posiadając żadnych przesłanek na temat danej gry, można założyć, że jest to gra „średnia”, której ocena wynosi  $(10 + 1)/2 = 5,5$  punktu. W praktyce punktacja BGG polega na dopisaniu każdej nowej grze stu<sup>23</sup> „próbnych” głosów fikcyjnych uczestników ankiety, z których każdy jest równy średniej ocenie wynoszącej 5,5 punktu.

W ten sposób nawet maksymalna ocena 10 punktów złożona przez pierwszego prawdziwego uczestnika ankiety podwyższy wynik BGG zaledwie do 5,54 punktów, ale też jedna jedynka tylko nieznacznie taką ocenę obniży. Gdyby 100 kolejnych graczy jednomyślnie przyznało ocenę 10, to ranking BGG gry będzie równy  $(100 \times 5,5 + 100 \times 10)/200 = 7,75$ , a w przypadku 1000 takich samych głosów ocena BGG wyniesie już 9,59. Zauważmy, że w miarę napływu kolejnych ocen rola początkowej setki „próbnych głosów Bayesa” w ogóle zanika, lecz ma ona spore znaczenie w przypadku gry nowej, ocenianej na początku przez niewielu graczy.

Omówiony system stosowany jest także w wielu innych rankingach internetowych<sup>24</sup> i wydaje się dobrze działać w praktyce. Poniżej wyciąg z listy gier planszowych z grudnia 2011 roku, na której gra *Dominion: Intrigue* oceniona przez wielu graczy, znajduje się wyżej niż *War of the Ring* oceniona nieco lepiej, ale przez znacznie mniejszą liczbę oceniających. Podobnie jak w innych przypadkach omówionych w tym rozdziale, i tym razem znaczącą rolę odgrywa arbitralnie wybierany element systemu, którym jest początkowa liczba „średnich ocen”.

Miejsce	Gra	Rocznik	Liczba ocen	Średnia ocena	Ranking BGG
1	<i>Twilight Struggle</i>	2005	9264	8,31	<b>8,201</b>
2	<i>Agricola</i>	2007	21500	8,25	<b>8,167</b>
3	<i>Dominion: Intrigue</i>	2009	9335	8,03	<b>7,896</b>
4	<i>War of the Ring</i>	2010	738	8,69	<b>7,634</b>
5	<i>Settlers of Catan</i>	1995	30525	7,51	<b>7,383</b>

## 11.5. Tenis, piłka nożna i szachy

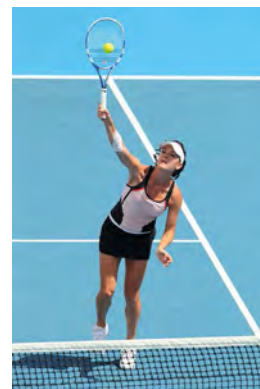
### Jak rozstawiać tenisistów w Wimbledonie?

Najstarszy i najbardziej prestiżowy **tenisowy** turniej odbywa się od roku 1877 na trawiastych kortach **Wimbledonu** – południowo-zachodniej dzielnicy Londynu. Zawody tradycyjnie rozgrywa się systemem pucharowym, w którym przegrywający mecz odpada z dalszej gry. W turnieju singla bierze udział 128 zawodników, więc do wygrania całych zawodów potrzeba siedmiu zwycięstw. Aby zmniejszyć prawdopodobieństwo, że dwóch najlepszych graczy spotka się już w początkowej fazie turnieju, przy losowaniu par stosuje się rozstawienie: **lider rankingu** z najwyższą liczbą punktów zdobytych w poprzednich zawodach może trafić na czwartą raketę turnieju dopiero w półfinale, a z graczem rozstawionym z numerem dwa zmierzy się nie wcześniej niż w finale.

Ale jak wylicza się pozycję zawodowego tenisisty w rankingu? Liczba punktów zdobytych za wygrany mecz zależy od rangi turnieju. Każdy uczestnik męskiego turnieju, który wchodzi w skład **Wielkiego Szlema** (w porządku chronologicznym: Australian Open w Melbourne, French Open na kortach Rollanda Garrosa, turniej w Wimbledonie oraz US Open na Flushing Meadows), otrzymuje 10 punktów za samo zakwalifikowanie się do turnieju, a następnie po 45, 90 i 180 punktów za przejście każdej z pierwszych trzech rund. Udział w ćwierćfinale daje 360 punktów, a wejście do pierwszej czwórki turnieju i start w półfinale wart jest 720 punktów. Uczestnik finału otrzymuje 1200 punktów, a zwycięzca całego turnieju – aż 2000. Tyle samo punktów otrzymuje najlepsza tenisistka turnieju, choć punktacja zawodów pań różni się nieznacznie od męskiej, i przykładowo udział w finale pań wyceniony jest na 1300 punktów.

Punktacja w innych turniejach, których prestiż zależy od wartości głównej nagrody, opiera się na tej samej zasadzie. Zwycięstwo w męskim turnieju rangi *World Tour Finals* to 1500 punktów, turnieju *World Tour Masters* – 1000 punktów, a turniejów mniejszej rangi nazywanych ATP 500 oraz ATP 250 – odpowiednio 500 lub 250 punktów. Udział w finale turnieju daje około 60% punktów przyznawanych za zwycięstwo, a dojście do półfinału i ćwierćfinału warte jest nieco więcej niż połowa oraz czwarta część tej liczby. Zwycięstwo w turnieju klasy *Challenger* daje od 80 do 125 punktów, a udział w nim – odpowiednio mniej.

Na całkowitą liczbę punktów rankingowych zawodnika w danym momencie składają się zsumowane wyniki co najwyżej 19 turniejów (panowie) lub 16 turniejów (panie) rozegranych w ciągu poprzedzających 52 tygodni. Na początku roku 2013 ranking najlepszych tenisistek



Agnieszka Radwańska, Australian Open 2011

świata otwierała Białorusinka Wiktorija Azarenka z 10 595 punktami przed Rosjanką Marią Szarapową – 10 045 punktów. Czwarte miejsce na świecie zajmowała polska tenisistka Agnieszka Radwańska z 7 425 punktami, a jej siostra Urszula zgromadziła 1 490 punktów, co dawało 31. pozycję. W klasyfikacji panów prowadził Serb Novak Djoković (12 920), a kolejne miejsca zajmowali Szwajcar Roger Federer (10 265) oraz Szkot Andy Murray (8 000). Po awansie do finału turnieju z serii *World Tour Masters* w Paryżu w październiku 2012 roku Polak Jerzy Janowicz z 1 299 punktami awansował na 26. miejsce w rankingu, a kolejny nasz rodak Łukasz Kubot zajmował 74. pozycję.

W ciągu ostatnich lat pewne szczegóły obliczania pozycji tenisistów w rankingu zmieniały się (na przykład punktacja poszczególnych zawodów, liczba turniejów w roku uwzględnianych w rankingu, punkty za starty w meczach drużynowych o Puchar Davisa, sposób ustalania rankingu dla zawodników pauzujących z powodu kontuzji), lecz ogólne zasady pozostawały niezmiennie. Wydaje się, że obecna postać rankingu pozwala obiektywnie opisać poziom każdego tenisisty i dobrze spełnia swe podstawowe zadanie: umożliwia dobre rozstawienie zawodników w kolejnym turnieju.

### Piłka nożna oraz ranking FIFA

Wartość meczu wygranego w turnieju tenisowym zależy bezpośrednio od rundy, w której mecz jest rozgrywany. Ale jak wycenić wartość zwycięstwa w danym pojedynku narodowych reprezentacji w **piłce nożnej**, skoro mecze o odmienną stawkę (towarzyskie, eliminacyjne) rozgrywane są z przeciwnikami o różnej klasie?

Aby określić siłę gry reprezentacji piłkarskich danego kraju, w grudniu 1992 roku **Międzynarodowa Federacja Piłki Nożnej (FIFA)** ustaliła zasady rankingu. W pierwotnej wersji uwzględniał on rezultaty każdej drużyny z okresu minionych ośmiu lat, lecz stosowana procedura prowadziła niekiedy do zaskakujących wyników. Głosy krytyków systemu nasiliły się, gdy w roku 1995 Norwegia znalazła się na drugim miejscu na świecie, oraz gdy w roku 2006 drużynę Stanów Zjednoczonych sklasyfikowano na miejscu czwartym, ku zaskoczeniu jej trenera i zawodników.

Po reformie systemu w roku 2006 pod uwagę bierze się wyniki wszystkich 208 reprezentacji narodowych za okres czterech lat wstecz. Liczba punktów *P* przyznawanych za każdy mecz wyraża się iloczynem czterech wskaźników

$$P = M \times I \times C \times T,$$

gdzie *M* to punkty za wynik (3 za zwycięstwo, 2 za zwycięstwo w rzutach karnych, 1 za remis lub przegraną w rzutach karnych, a 0 za porażkę). Współczynnik *I* opisuje rangę meczu i wynosi 1 dla meczu towarzyskiego,

2,5 dla meczu eliminacji mistrzostw świata lub kontynentu, 3 za mecze w finałach mistrzostw kontynentu oraz 4 za mecze finałów mistrzostw świata. Współczynnik  $C$  opisuje względną siłę rozgrywek na różnych kontynentach i powoli zmienia się w czasie ( $C = 1$  dla Europy i Ameryki Południowej; od roku 2010 dla Ameryki Północnej, Środkowej i Karaibów współczynnik  $C = 0,88$ , dla Afryki  $C = 0,86$ , a dla Azji i Oceanii  $C = 0,85$ ). Ostatni czynnik opisuje siłę rywala wzorem  $T = (200 - p)/100$ , gdzie  $p$  oznacza pozycję przeciwnika w rankingu. Przykładowo, gdy przeciwnikiem jest ósma drużyna świata, to wartość tego czynnika wynosi  $T = (200 - 8)/100 = 1,92$ . Pewnym wyjątkiem od tej reguły jest współczynnik dla pierwszej drużyny w rankingu, dla której przyjmuje się  $T = 2$  oraz dla słabszych zespołów zajmujących miejsca 150 lub niżej, dla których  $T = 0,5$ . Sumując punkty za każdy oficjalny mecz drużyny narodowej za okres czterech lat, wylicza się punkty rankingowe danej reprezentacji.

Stosując powyższe reguły, otrzymuje się dosyć rozsądne wyniki, a otrzymane uporządkowanie drużyn nie budzi już dużych wątpliwości. Na pierwszym miejscu rankingu FIFA przez ponad 10 lat znajdowała się reprezentacja Brazylii, a przez pewien czas liderami listy były też Francja, Włochy i Argentyna. W grudniu 2012 roku w rankingu prowadziła Hiszpania (1606 pkt.), przed Niemcami (1437) i Argentyzną (1290). W marcu tego roku reprezentacja Polski zajmowała 75. miejsce, jak dotąd najgorsze w historii, lecz po kilku remisowych meczach z wyżej notowanymi rywalami awansowała w pod koniec roku 2012 na 55. pozycję z dorobkiem 553 punktów. W tym samym czasie analogiczny ranking drużyn kobiecych otwierały piłkarki Stanów Zjednoczonych przed reprezentacjami Niemiec i Japonii, a Polki zajmowały 31. miejsce na świecie.

Warto zauważyć, że przyznanie Polsce organizacji mistrzostw Europy w roku 2012 wpłynęło negatywnie na ranking polskich piłkarzy: jako gospodarz zawodów EURO 2012 drużyna polska nie musiała brać udziału w eliminacjach do tego turnieju, przez co w ostatnich latach rozgrywała jedynie niżej punktowane mecze towarzyskie. Podobnie organizator kolejnych mistrzostw świata, Brazylii, spadła w grudniu 2012 roku na najniższe w historii, 18. miejsce w rankingu<sup>25</sup>. Najwyższe, 16. miejsce w rankingu, reprezentacja Polski zajmowała we wrześniu 2007 roku. Można żałować, że rankingowi FIFA nie obliczano już w latach 70. XX wieku, gdyż niezapomniana drużyna Kazimierza Górskiego (złoto olimpijskie w roku 1972, trzecie miejsce na Mundialu w roku 1974) z całą pewnością byłaby sklasyfikowana w pierwszej dziesiątce na świecie.

## System Elo do oceny poziomu gry szachisty

Punktowa wartość wygranego meczu reprezentacji krajów w piłce nożnej zależy od miejsca, jakie zajmuje przeciwnik w rankingu FIFA. Także



Polska–Gruzja 1:0, Lubin, 2011

25) Polecamy analizę Marka M. Kamińskiego, *Jak silna jest polska piłka nożna? Paradoks „gospodarza turnieju” w rankingu FIFA*, zamieszczoną w piśmie „Decyzje” 2012, nr 17, s. 29-45.



Arpad Elo (1903–1992) – twórca rankingu szachistów

26) Zmiana punktów gracza  $A$ , który ma przed partią  $R_A$  punktów rankingowych, po pojedynku z szachistą  $B$ , który ma ich  $R_B$ , wynosi  $D = K \times [w - 1/(1+10^d)]$ , gdzie wynik partii  $w = 1$ , gdy zwycięży gracz  $A$ ,  $w = 1/2$  w sytuacji remisu i  $w = 0$  w przypadku porażki  $A$ , zaś wykładnik  $X = (R_B - R_A)/c$ . W oryginalnym systemie Elo przyjęto  $c = 400$ , a wartość  $K$  była równa 10. Obecnie wartość tej stałej w rankingu FIDE zależy od poziomu graczy:  $K = 30$  dla graczy rozpoczynających karierę,  $K = 15$  dla szachistów o liczbie punktów poniżej 2400, a  $K = 10$  dla mistrzów szachowych z wyższą pozycją w rankingu.

27) Zwycięstwo Carlsena w turnieju *London Chess Classic* w grudniu 2012 roku pozwoliło temu 22-letniemu Norwegowi przekroczyć granicę 2860 punktów i wyjść na czoło rankingu szachistów wszech czasów.

w szachach używa się punktów rankingowych do określenia relatywnej siły gry zawodników, a po każdej partii pozycja zwycięzcy w rankingu rośnie, zwyciężonego zaś – maleje. W przypadku remisu nieco punktów zyskuje gracz niżej notowany. O liczbie punktów rankingowych  $D$  za zwycięstwo w danej partii szachowej decyduje różnica pomiędzy liczbami punktów rankingowych obu przeciwników. **Międzynarodowa Federacja Szachowa (FIDE)** stosuje system rankingu opracowany w roku 1960 przez amerykańskiego fizyka (i szachistę!) pochodzenia węgierskiego, **Arpada Elo**.

**System Elo**, tylko nieznacznie modyfikowany w ostatnich pięćdziesięciu latach, bazuje na założeniu, że wynik partii szachów można traktować jako zmienną losową o rozkładzie normalnym. Metody statystyczne pozwalają wyprowadzić jawny wzór<sup>26</sup> na liczbę punktów zdobytych w partii, który co prawda wymaga umiejętności potęgowania, lecz nietrudno stosować go za pomocą kalkulatora.

Ranking szachisty rozpoczynającego karierę wyznacza się na podstawie kilku partii rozegranych z graczami już posiadającymi punkty rankingowe. Liczba 1000 punktów stanowi ograniczenie dolne, więc każdy szachista obecny na liście FIDE ma nie mniej niż tysiąc punktów. W miarę wzrostu umiejętności szachiści mierzą się z coraz lepszymi przeciwnikami, którzy wcześniej stoczyli już wiele partii, a ich siła gry jest dobrze opisana przez punkty rankingowe. W ten sposób każda notowana partia szachowa prowadzi do pewnych zmian w rankingu i pozwala na dokładniejsze określenie aktualnej siły gry danego gracza.

Niezależnie od punktów rankingowych, każdy szachista, biorąc udział w odpowiednich turniejach, zdobywa normy na kategorie sportowe (od najniższej, piątej kategorii, do najwyższej – pierwszej). Wygrywając w turniejach z szachistami pierwszej kategorii, zdobywać można tytuły mistrzowskie. O ile ranking Elo dość precyzyjnie określa *aktualny* poziom gry szachisty, to zdobyta kategoria lub tytuł określa *największy* sukces danego gracza. Zdobycie odpowiedniej liczby punktów jest warunkiem koniecznym do uzyskania danego tytułu. Przykładowo, należy zdobyć ponad 2300 punktów rankingowych, aby mieć szansę na tytuł **mistrza FIDE**. Najlepsi szachiści z tytułem **mistrza międzynarodowego** przeważnie zgromadzili ponad 2400 punktów Elo, a gracze z dorobkiem powyżej 2600 punktów zdobyli już tytuł **arcymistrza**. Wśród pań tytuł **arcymistrzyni** otrzymują szachistki, które przekroczą granicę 2300 punktów.

Według danych z końca 2012 roku najwyższą w historii liczbę punktów według rankingu FIDE uzyskał w roku 1997 Garri Kasparow (Rosja), który zgromadził ich wówczas 2851. Drugi na liście był Magnus Carlsen<sup>27</sup> (Norwegia), który miał 2848 punktów, trzeci – Lewron Aronian (Armenia), który uzyskał ich 2825. Na 37. miejscu listy znajdował się Radosław Wojtaszek (Polska) z 2734 punktami. Wśród pań najwyższą liczbę punktów w historii

zgrupowała Judit Polgár (Węgry) – 2735, a Polka Monika Soćko, z liczbą 2505 osiągniętą w roku 2008, zajmowała 33. miejsce. Warto podkreślić, że w miarę upływu czasu i zwiększania się liczby klasyfikowanych szachistów oraz liczby rejestrowanych partii ranking najlepszych graczy stopniowo rośnie. Przykładowo, w październiku roku 2002 średnia liczba punktów pierwszej setki szachistów świata wynosiła 2649, a 10 lat później – już 2704 punkty.

W roku 1997 grupa entuzjastów statystyki i footballu zastosowała analogiczny ranking (pod nazwą *World Football Elo*) do piłki nożnej. Okazuje się, że daje on odmienne wyniki od rankingu FIFA: Brazylia na przykład plasowała się w grudniu 2012 roku w tym rankingu na drugim miejscu, podczas gdy w rankingu FIFA, jak już wspominaliśmy, na 18. Piłkarskie Elo można też zastosować do przeszłości. Okazuje się, że polska reprezentacja między czerwcem 1974 roku a październikiem 1975 roku plasowała się w tym rankingu stale na drugim lub trzecim miejscu na świecie, a przez kilkanaście lat nie opuszczała pierwszej dziesiątki! Trudno o lepszy dowód, że byliśmy niegdyś piłkarską potęgą.

Szachowy ranking Elo – pierwszy system oparty na rozumowaniu statystycznym – został dobrze skonstruowany i służy szachistom już 50 lat. Ale system ten ma też słabsze strony. Bazuje on na najczęściej używanym w statystyce rozkładzie normalnym, podczas gdy niedawne analizy pokazują, że dane empiryczne są w tym przypadku lepiej opisywane rozkładem logistycznym. Ponadto system Elo nie uwzględnia dokładności, z jaką określony jest ranking każdego gracza. Rozważmy partię szachistów *A* i *B*, którzy obaj mieli po 1700 punktów. Partię wygrywa gracz *A*. Według aktualnego rankingu Elo zwycięzca partii zyska 15 punktów, które straci szachista *B*. Załóżmy jednak, że gracz *A* nie rozegrał żadnej partii w ciągu ostatniego roku, podczas gdy *B* gra regularnie. W tej sytuacji liczba 1700 dobrze określa siłę gry szachisty *B*, lecz jest mniej wiarygodna przy opisie poziomu gracza *A*. Sądząc po wyniku partii, z dużym prawdopodobieństwem rzeczywista siła gry szachisty *A* jest obecnie większa niż 1700. Dlatego bardziej sprawiedliwe byłoby przyznanie mu za to zwycięstwo więcej niż 15 punktów, a gracz *B* na partii z być może znacznie silniejszym przeciwnikiem powinien stracić mniej punktów niż 15.

Taka argumentacja doprowadziła amerykańskiego statystyka **Marka Glickmana** do stworzenia w roku 1995 nowego systemu opartego na statystycznym podejściu Bayesa, w którym każdemu szachiście przypisuje się zarówno liczbę punktów rankingowych, jak i wielkość rozrzutu, z jakim jego pozycja w rankingu jest określona. Liczba punktów zyskiwana lub tracona po rozegranej partii zależy od jej wyniku oraz wszystkich czterech parametrów obu graczy, a dokładny wzór nietrudno jest odnaleźć

w internecie. Stosując swój system do analizy dawnych partii, Glickman ułożył listę szachistów wszech czasów, na czele której stoją byli mistrzowie świata: Emanuel Lasker, a za nim José Capablanca, Robert Fischer, Aleksander Alechin i Garri Kasparow. Taki system można stosować także w innych dziedzinach, a listę otrzymaną dla tenisistów czynnych w roku 1995 otwiera André Agassi, podczas gdy kolejne miejsca zajmują Pete Sampras i Thomas Muster.

Modyfikacja rankingu Glickmana z roku 2000, zwana **Glicko-2**, jest już stosowana przez Amerykańską Federację Szachową. Wobec olbrzymiego wzrostu możliwości obliczeniowych komputerów, przygotowanie programów wyliczających pozycje rankingowe wedle nieco bardziej skomplikowanych wzorów nie nastrocza kłopotów. Dlatego można się spodziewać dalszego rozwoju rankingów szachowych oraz ich coraz powszechniejszego wykorzystania.

## 11.6. Jak wybrać zwycięzcę konkursu?

Przedstawione przykłady z różnych dziedzin życia pokazują, że nie ma idealnego systemu wyłaniania zwycięzcy w konkursie sztuki czy w zawodach w sportach niewymiernych. Przykład pechowego zapaśnika z Bułgarii, któremu koło nosa przeszedł olimpijski medal, pokazuje, że dobry system wyłaniania zwycięzcy zawodów powinien uwzględniać możliwość wystąpienia także mało prawdopodobnych zdarzeń. Tak jak twierdzenie matematyczne jest prawdziwe przy każdym zestawie danych spełniających zadane założenia, tak przyjęty system rozgrywania zawodów powinien prowadzić do rozsądnych rezultatów przy dowolnym układzie wyników poszczególnych walk i meczów.

Nawet najbardziej wymyślny system oceny nie będzie jednak dawał rozsądnych wyników, jeżeli niektórym członkom gremium oceniającego zabraknie kompetencji, bezstronności i dobrej woli. Z drugiej strony, jeżeli oceniający mają do siebie (uzasadnione!) zaufanie, to nawet prosty, mało sformalizowany system oceniania w konkursie może przynieść dobre rezultaty.

Dlatego wybierając system punktacji i opracowując regulamin konkursu, warto stosować proste rozwiązania. Bardzo skomplikowany czy wymyślny system punktacji i tak nie zabezpieczy organizatorów przed wszystkimi możliwymi nadużyciami, a skomplikuje życie jury i zwiększy prawdopodobieństwo pomyłek i błędów. Większą rękojmią rzetelnych rezultatów będzie właściwy wybór składu jury, którego kompetentni członkowie wzajemnie sobie ufają.

Zadaniem każdego z sędziów jest przygotowanie obiektywnej i wiarygodnej oceny każdego z uczestników zawodów. W konkursach i zawodach organizowanych w formie dwóch (lub więcej) etapów, do swoistej konkurencji dopuścić można także... sędziów. Nie poprzez zamianę ról z uczestnikami, lecz poprzez bezpośrednią ocenę jakości ich pracy. Przykładowo, w pierwszym etapie zawodów zliczać można, ile razy oceny każdego sędziego są traktowane jako skrajne i nie wchodzi do obliczania średniej oceny zawodnika. Do udziału w finałowym etapie zawodów kwalifikowaliby się z jednej strony najlepsi uczestnicy konkursu, a z drugiej strony ci sędziowie, którzy popełniliby najmniej pomyłek w pierwszej turze zawodów.

Analizując wyniki każdego rankingu w dowolnej dziedzinie, warto pamiętać, że końcowa kolejność w sposób istotny zależy od wyboru kryteriów i wag, z którymi zostały one uwzględnione. Dlatego wynikami

rankingów nie należy posługiwać się bezkrytycznie. Nawet najlepiej dopracowany ranking rzadko pozwala na wyłonienie bezapelacyjnego zwycięzcy danej konkurencji, ale ułatwia rozróżnienie między najlepszymi, średnimi i słabszymi kandydatami. Jak pokazuje zadziwiająco wysoka pozycja Norwegii w piłkarskim rankingu FIFA w roku 1995, pozornie rozsądna konstrukcja rankingu może też niekiedy prowadzić do paradoksalnych rezultatów. Każdy system rankingowy warto jest więc przetestować na rzeczywistych danych oraz dokładnie zbadać, jak zmiany parametrów systemu wpływają na otrzymywane wyniki.