

СТАНОВИЩЕ

за дисертационния труд на Антон Дончев

„Вероятностни подходи към потвърждението. Ролята на приорните вероятности в правенето на извод в полза на най-доброто обяснение“

от доц. д-р Росен Люцканов (ИИОЗ-БАН)

Без съмнение „Вероятностни подходи към потвърждението“ е безупречно осъществен дисертационен труд. В него има всичко, което трябва да бъде налице: ясно очертана предметна област, прецизно формулирана теза, адекватни реконструкции на съвременните подходи, богата и представителна литература, оригинални хрумвания, несъмнени приносни моменти. Всъщност, текст като този до голяма степен е прецедент, защото това е дисертация, която би могла да бъде написана и защитена в кой да е добър философски факултет по света. Като се има предвид този факт, авторът на становището има две възможности: да се ограничи с регистрирането на многобройните предимства на текста, или да опита да предложи начини той да бъде развит и подобрен (колкото и трудно да е това). Избирам втората възможност, съзнавайки ясно, че предложенията ми спадат към една от следните две категории: (1) „това би могло да се каже и иначе“ и (2) „тук би могло да се добави още нещо“. Нито един от двата типа коментари не е критика към текста в настоящия му вид, правя ги единствено с оглед на възможността това изследване да бъде публикувано, при това в престижно издателство. Както посочва Дончев, има поне два различни и привидно несъвместими подхода към потвърждаването на едно твърдение - „от една страна, чрез нови емпирични свидетелства, а от друга – чрез качеството на обяснението, което въпросното твърдение предоставя“ (с. 7). Тези подходи са в основата съответно на бейсианската теория на потвърждението (ВСТ) и извода в полза на най-доброто обяснение (IBE). Според Дончев самото наличие на тези алтернативни подходи „отваря въпроса за тяхната съвместимост“, още повече, че те поне на пръв поглед биха могли да се допълват взаимно, компенсирайки своите недостатъци - „ВСТ разполага с добре развит математически механизъм, основан на теория на вероятностите, но в него липсват инструменти за определянето на някои ключови параметри. IBE има нуждата от точна формална експликация, каквато бейсианският механизъм може да предложи, като IBE от своя страна изглежда може да предостави необходимите инструменти, които липсват в ВСТ“ (с. 7). Всъщност, основният проблем на IBE е свързан с въпроса как да разпознаем „най-доброто“ обяснение. Разбира се, това не е единствения проблем – други проблеми са свързани с асиметричността на абдуктивните изводи (с. 64) и невъзможността въз основа на принципни съображения да елиминираме така наречените „спекулативни абдукции“ (с. 75). Решението, което Дончев предлага, е „понятието „най-добро обяснение“ да бъде дефинирано по следния начин: „най-доброто обяснение от множество съревноваващи се обяснения е онова, което има най-голяма обяснителна сила““ (с. 8). Съответно, централната теза на дисертационния труд е, че „понятието „най-добро обяснение“, което е ключово за теорията на потвърждението, основаваща се на „извод в полза на най-доброто обяснение“ (IBE), може да бъде получи адекватна вероятностна експликация чрез бейсианска мярка на обяснителната сила, при условие, че последната отчита влиянието на приорните вероятности“ (с. 8). С цел аргументиране на тази централна теза, изложението се разгръща в пет глави: увод, който поставя проблемния контекст на изследването, въведение в съвременната теория на потвърждението, преглед на опитите за съвместяване на ВСТ и IBE, анализ на ролята на приорните вероятности в бейсианската експликация на понятието за „най-добро обяснение“ и обобщено представяне на резултатите от изследването, набеязващо редица отворени въпроси, представляващи интерес за по-нататъшна работа (с. 11-12). Още тук бих искал да отбележа, че резултат от изследването е не съвместяване на ВСТ и IBE, а по-скоро редуциране на едното до другото. Наистина, Дончев свежда „най-добро обяснение“ до „обяснение с най-голяма обяснителна сила“, а „обяснение с най-голяма обяснителна сила“ до „обяснение с най-висока степен на потвърждение“ (с. 146). Какви са резултатите от тази редукция става ясно като сравним бейсианските мерки на потвърждение (с. 40) и директните

бейсиански мерки на обяснителна сила (с. 118): E_1 се получава от C_1 , E_4 се получава от C_3 , E_5 се получава от C_6 и E_6 се получава от C_7 като просто разменим местата на e и h в съответната формула. Иначе казано, оказва се, че обяснителната сила на h спрямо e съвпада със степента на потвърждение на h относно e . Ето защо не съм убеден, че привържениците на ИВЕ биха се зарадвали на брака по сметка с ВСТ, който Дончев им предлага. Освен това, струва ми се, че тук би могъл да възникне сериозен проблем (вж. приложението в края на текста).

Изходна точка на изложението е разграничаването на два основни модела на потвърждение: логически и вероятностен (с. 18). Прегледът на стандартните версии на хипотетико-дедуктивизма с техните неотстранени проблеми (проблемът с нерелевантната конюнкция, самопотвърдимостта, парадоксите на концептуалната разлика и този на гарваните, както и новата загадка на индукцията на Гудман) (с. 23-27) водят Дончев до извода, че в резултат от тези проблеми „в края на ХХ фокусът на вниманието в теория на потвърдението се отмества категорично към вероятностните модели на потвърждение“ (с. 29). По повод на това бих искал да направя два коментара: (1) логическият подход не просто е изместен от вероятностния, той по-скоро е ликвидиран с чисто логически средства, което демонстрира колко силен инструмент е логиката, дори когато я насочим срещу самата нея (разбира се, тук имам предвид диагоналия аргумент на Пътнам срещу „решението“ на Карнап на проблема за индуктивната логика); (2) не съм съгласен, че изместването е „категорично“; разбира се, бейсианството е изключително модерен подход, но формалната теория на обучението (formal learning theory) също има свои привърженици (в тази връзка бих препоръчал прекрасната книжка Earman, J. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press; последната глава сравнява двата подхода), както и следните два ключови текста: Kelly, K. 1996. *The Logic of Reliable Inquiry*. Oxford: Oxford University Press и Martin, E. and D. Osherson. 1998. *Elements of Scientific Inquiry*. Cambridge, MA: MIT Press.

Втора глава предлага стегнато изложение на стандартната версия на теорията на вероятностите (тук бих искал само да отбележа, че на с. 30 трябва да заменим „естествени“ с „реални числа“). Съвсем коректно е посочено също, че съществуват и други опити за нейното аксиоматизиране (с. 34). Според мен изложението би спечелило от дори кратък преглед на такива алтернативи – например на най-значимото обобщение на субективния бейсиански подход (теорията на Демпстър-Шейфър, най-вече по повод на ключовото за нея разграничение между потвърденост и правдоподобност), или на качествените подходи към анализирането на вероятности, които тръгват още от де Финети (при тях не оценяваме вероятности, а само ги сравняваме; това е важно по повод връзката с ИВЕ, тъй като при нея също се използват качествени мерки). В тази връзка бих препоръчал съответно Yager, R. and L. Liu (eds.). 2008. *Classic works of the Dempster-Shafer theory of belief functions*. Berlin: Springer. и Krantz, D., D. Luce, P. Suppes, and A. Tversky. 1971. *Foundations of Measurement*, vol. I. New York Academic Press (по-точно гл. V: Probability Representations). Всичко това на пръв поглед отива встрани от основната линия на текста, но според мен това впечатление е погрешно: допускането, че вероятностите се измерват чрез реални числа и съответно могат да се оценяват с произволна степен на точност е теоретична идеализация, която поне привидно проблематизира приложенията на теорията за анализиране на реални ситуации на взимане на решения от страна на агенти с ограничена рационалност (дори те да са учени).

Изложението продължава с изключително стегнат анализ на вътрешните разделения в рамките на бейсианството – между субективисти и обективисти. Разликата между тях се проявява най-вече по повод на основната тема на текста – приорните вероятности и начините за тяхното определяне (с. 38). Разгледани са също и различни проблеми на бейсианския подход – несигурността на свидетелствата (с. 41), трудностите със старите свидетелства (с. 42), логическото всезнание (с. 42) и проблемът за приорите (с. 43). Според Дончев „едно от популярните предложения за съвместяване на двата подхода към потвърдението – бейсианската теория и извод в полза на най-доброто обяснение (ИВЕ), е последното да играе ролята на механизъм за определянето на приорните вероятности“ (с. 43), така че последният проблем се оказва особено важен. По повод на горния списък бих искал да отбележа няколко

допълнителни аспекта: (1) епистемичната логика отдавна се бори с проблема за логическото всезнание, така че някои изведени от нея поуки вероятно биха свършили работа и в този контекст; (2) склонността да пропускаме наглед невинната буквичка „к“ когато изписваме съответните формули скрива този проблем, като създава предпоставки да забравим, че целият „познавателен фон“ следва да бъде отчитан във всеки отделен случай – едно твърде нереалистично допускане; (3) според мен основният проблем на бейсианството не присъства в горния списък: той е свързан с това, че то предвижда единствено преразглеждане на приорната вероятност на хипотезата, по-точно нейно заместване с постериорната условна вероятност; работата е там обаче, че в редица случаи изглежда рационално да преразглеждаме както условните вероятности, така и самите приорни вероятности (както посочват някои автори, едно неуспешно предсказание може да се дължи и на това, че сме определили неправилно приорната вероятност и съответно се налага самата тя да бъде преразгледана; при това положение ще ни трябва не еднократно прилагане на правилото на Бейс, а нещо съвсем различно); (4) едно от безспорните предимства на класическия бейсиански подход е свързано с това, че прилагането му ни защитава от холандски залози; от друга страна, това не е единствения начин да ги избегнем – друг пример е принципът за максимална ентропия, още повече, че можем да разглеждаме теоремата на Бейс като негов частен случай. Ето защо, този тип аргументи в полза на бейсианството не работят.

По-нататък изложението продължава с набелязване на приликите и разликите между абдукция и индукция (с. 51); това е важно, защото бейсианската теория на потвърдението е свързана с индукцията, а извода в полза на най-доброто обяснение – с абдукцията. В тази връзка основно значение добива въпросът дали изобщо има и може да има такова нещо като „абдуктивен метод“. Един от най-известните критици на абдукцията е ван Фраасен, който формулира три влиятелни аргумента срещу нея (с. 53-54). Изводът, който той прави, гласи, че „обяснителната стойност на една научна теория (или хипотеза) може да бъде важна за нейното приемане или отхвърляне, но не може да бъде важна за нейното потвърждение или опровержение“ (с. 54). Този скептически извод относно възможността за обединяване на ВСТ и ИВЕ до голяма степен бива разколебан по-нататък, макар и с цената на заявлението, че отклоненията от правилото на Бейс не са задължително симптом за ирационалност (с. 57). Текстът предлага също изключително интересен и компресиран анализ на отношението между обяснение и разбиране. Представени са различни съвременни концепции за разбирането: като предоставяне на основания за формиране на убеждение, като свеждане на непознатото към познатото, като унификация на различни по вид феномени, като демонстриране на необходимост и посочване на причини (с. 80-84).

Трета глава продължава с представяне на различните компатибилистки подходи, които целят да покажат, че ВСТ и ИВЕ не само могат, а трябва да бъдат съчетани. Съответно, те трябва да преодолеят предизвикателствата на инкомпатибилизма, много от които са свързани с критики срещу ИВЕ (с. 92-94). Въз основа на тъкмо такива критики, Салмън стига до извода, че „потвърдението е логически независимо от обяснението“ (с. 95), това е поредния голям препъни-камък по пътя към обосноваването на централната теза след критиките на ван Фраасен. Тук на помощ са призовани влиятелни компатибилистки подходи като тези на Хартман и Вебер (с. 98), Окаша (с. 101), Хендерсън (с. 102-103), Липтън (с. 104-105), Уайсберг (с. 114-115) и Псилос (115-116). Резултатите от този пространен преглед са резюмирани от Дончев по следния донякъде обезкуражаващ начин: „съществуващите за момента опити за интеграция на ВСТ и ИВЕ се провалят. Нито едно от разгледаните по-горе предложения не е достатъчно убедително или пък ясно, за да реши проблема със съвместимостта на ВСТ и ИВЕ“ (с. 116). Това обаче има положителен смисъл, тъй като доказва, че дисертацията си поставя за задача да реши един наистина нетривиален проблем.

Така стигаме до трети параграф на трета глава – ключово място в хода на изложението, където лесно можем да забележим оригинални идеи с приносен характер. Основната задача тук е да се провери как е възможно измерването на обяснителната сила на дадена хипотеза (с. 116). Почти веднага е набелязан инструмент, който изглежда подходящ за целта -

„очакваността на свидетелствата“, която при това кореспондира на интуицията, че „доброто обяснение прави иначе неочакваните и изненадващи свидетелства очаквани“. Ето защо, тя би следвало да играе съществена роля в директните мерки на обяснителната сила (с. 117). Разгледан е също въпросът за дескриптивната адекватност на мерките от този тип, която може да бъде заплашена от това, че мерките биват прилагани към дейци като хората, които рядко (или почти никога) не демонстрират онази перфектна рационалност, която бейсианските модели изискват от тях (с. 119). Така или иначе, въпросът коя от предложените мерки е най-добра все още остава отворен, най-вече защото всеки от многото автори защитава своето собствено предложение по въпроса (с. 120). Всички те са критикувани (с основание) от Кларк Глимур, който посочва, че нито една от предложените мерки няма ясно специфициран обхват на приложение, а освен това те трудно са приложими на практика, тъй като съответните параметри рядко са изчислими по еднозначен начин (с. 122). Друг проблем е свързан с въпроса как точно трябва да приписваме стойност на субективните вероятности, които участват в съответните формули (с. 123). Накрая, „някои от мерките приписват положителна обяснителна сила на две следствия едно спрямо друго“ (с. 123). Това е проблем, който многократно е засвидетелстван в различни теории за научното обяснение.

В заключение Дончев посочва, че „всички директни мерки на обяснителната сила не отчитат приорните вероятности на сравняваните хипотези“ (с. 124, 127) и може би това е причината за всичките им неблагоприятия. За мен интересна е връзката между този проблем и „пренебрегването на базовата честота“ (с. 130); вероятно на тази тема си струва да се отдели повече внимание. Парадоксално, тази връзка би могла да се използва като аргумент в подкрепа на бейсианските мерки – ако прилагането им води до грешки, които отговарят на грешки, които самите ние допускаме в своите разсъждения, значи може би те все пак са дескриптивно адекватни, защото правдиво описват поне някои от нашите разсъждения.

Четвърта глава започва с един от задължителните елементи на всеки монографичен текст по философия на науката – анализ на конкретни примери от нейната история. Те целят да демонстрират ролята на приорните вероятности по отношение на оценката на правдоподобността на една хипотеза от страна на съответната научна общност. Според мен обаче, примерите реално показват, че върху оценката на дадена хипотеза влияние оказват не само епистемичните стандарти в науката и принципите на бейсианската епистемология, а също и чисто социологически фактори – в крайна сметка, проблемите на Вегенер и неговата хипотеза за континенталния дрейф изглежда се дължат на това, че той е метеоролог, а не геофизик, съответно теорията му е построена по начин, който не изглежда правдоподобен за специалистите в областта (с. 138). Другият анализиран пример – с още неоткритата „планета девет“ сякаш показва същото: хипотезата на Батигин и Браун изглежда за предпочитане пред своите конкуренти (с. 142), защото е построена по аналогия с тази на Льоверие, който предсказва съществуването на осмата планета – Нептун. Това е типичен за науката (и не само за нея) консерватизъм: ако даден подход е проработил преди, това е силен аргумент да го изпробваме отново, преди да търсим нови и донякъде „екзотични“ решения.

Най-важната теза в цялата дисертация според мен е изказана на с. 145: „ако приемем за адекватна експликация на понятието „най-добро обяснение“ някаква бейсианска мярка на обяснителната сила, която взема предвид и приорните вероятности, бихме избегнали контраинтуитивното заключение, че съревноваващите се обяснителни хипотези имат еднаква обяснителна сила“. От тази теза изхожда разработването на метод за изследване на условията за съвместимост между ВСТ и ИВЕ, който е скициран на следващите страници и според Дончев може безпроблемно да бъде приложен към онези интерпретации на ИВЕ, които използват вероятностна експликация на понятието „най-добро обяснение“ (с. 148). Приложението на този метод е илюстрирано с една от предложените мерки; според мен би имало голямо приносно значение, ако той бъде приложен и към останалите и в резултат се формулират общи изводи или евентуално се набележат други мерки, които до този момент не са известни. Това е идея, до която очевидно самият Дончев е стигнал (с. 152-153). Така или иначе, той стига до заключението, че „както инкомпатибилизъмът, така и компатибилизъмът, за

момента се провалят в решаването на основния проблем, който разглеждат. Нито една от двете позиции не успява убедително да покаже, че двата подхода към потвърждението са напълно съвместими или напълно несъвместими“ (с. 150). Този „отворен финал“ адекватно обобщава резултатите от дискусиата и същевременно обещава плодотворно развитие на набелязаната от дисертанта изследователска програма. Имайки предвид това обстоятелство без колебание бих поставил изключително висока оценка на дисертационния труд и убедено ще гласувам за това на Антон Дончев да бъде присъдена научно-образователната степен „доктор“ по направление 2.3 (философия). Несъмнено той е един от най-талантливите млади учени в България днес и бъдещ утвърден философ на науката.

07.06.2018,
София

/Р. Люцканов/

Приложение: Случаят с бай Иван и директните мерки за обяснителна сила

Ако най-добро е това обяснение, което посочва хипотеза с най-висока степен на потвърждение, то в много случаи онези обяснения, които изхождат от наблюдавани събития и на които съответно приписваме вероятност единица, ще бъдат неоснователно привилегирани спрямо останалите, които имат по-малка вероятност. За да се убедим в това, нека разгледаме следния очевидно скалъпен пример: Бай Иван живее в малко селце в България. Една лятна вечер го виждаме да се прибира към къщи с олюляваща се походка и ставаме свидетели как си удря главата в уличен стълб. Малко по-късно чуваме от негов познат, че е починал от мозъчен кръвоизлив. Имаме две алтернативни хипотези: (1) кръвоизливът е причинен от удара в стълба; (2) кръвоизливът е причинен от прекомерна консумация на метилов алкохол (разбира се, списъкът не е изчерпателен: бай Иван би могъл да е получил инсулт). Кое обяснение ще бъде по-добро? Видели сме бай Иван да си удря главата (съответно, присвояваме на това събитие вероятност $P(h_1) = 1$, тъй като нямаме основания да допуснем, че самите ние сме халюцинирали в този случай), но не и да консумира алкохол (ето защо, присвояваме на това събитие вероятност $P(h_2) = 1/2$, съгласно с повелята когато нямаме нужната информация да присвояваме равни вероятности на разглежданите алтернативи). Оценяваме тези хипотези от гледна точка на допускането, че е налице мозъчен кръвоизлив (e), на което също приписваме някаква вероятност (конкретната стойност не е от значение). По-нататък приемаме, че 90% от мозъчните кръвоизливи са причинени от механични травми, а едва 1% - от отравяне с метилов алкохол (тъй като за щастие случаите на такова отравяне са изключително редки). Разбира се, тук стои въпросът коя е референтната група, по отношение на която оценяваме честотите – самият бай Иван и историята на неговите неблагоприятия, възрастните хора, живеещи в малки населени места или цялата популация (разбира се, този въпрос е свързан с парадокса на Симпсън). Така или иначе, използвайки честотната интерпретация на вероятностите и теоремата на Бейс ще стигнем до извода, че условната вероятност $P(e|h_1)$ е много по-голяма от $P(e|h_2)$. Съответно, според повечето от предложените директни мерки за обяснителна сила (по-точно, E_1 , E_3 , E_4 и вероятно E_6) ще се окаже, че обяснението с механична травма е по-добро от това с алкохолно отравяне. Дали обаче наистина това е така? Всичко зависи от това, как ще опишем наблюдаваната поредица от събития, иначе казано, как ще разделим пространството на събитията – едно е „Бай Иван си удари главата в стълб“, съвсем друго „Бай Иван си удари главата в стълб, прибирайки се с олюляваща се походка към къщи“. Ето защо, според мен, обяснителната сила на една хипотеза не може да се оценява иначе, освен на фона на начина, по който се парцелира логическото пространство на всички възможни събития. Разбира се, Дончев осъзнава прекрасно това и го е посочил еднозначно в текста си (с. 115). От това следва обаче, че отчитането на приоритните вероятности съвсем не е панацея (или пък не?).