

НОВ БЪЛГАРСКИ УНИВЕРСИТЕТ
ДЕПАРТАМЕНТ „ФИЛОСОФИЯ И СОЦИОЛОГИЯ“



АНТОН ДОНЧЕВ

**ВЕРОЯТНОСТНИ ПОДХОДИ КЪМ ПОТВЪРЖДЕНИЕТО.
РОЛЯТА НА ПРИОРНИТЕ ВЕРОЯТНОСТИ В ПРАВЕНЕТО НА
ИЗВОД В ПОЛЗА НА НАЙ-ДОБРОТО ОБЯСНЕНИЕ**

АВТОРЕФЕРАТ

на дисертация за присъждане на образователната и научна степен „доктор“
в професионално направление 2.3 Философия

НАУЧЕН РЪКОВОДИТЕЛ: ДОЦ. Д-Р ЛИЛИЯ ГУРОВА

София, 2018

Съдържание

I. Увод.....	3
1. Централна теза.....	4
2. Основни цели на изследването	5
3. Метод на изследването	5
4. Структура на изложението	6
II. Теория на потвърдението	7
1. Основи на теорията на потвърдението	7
2. Кратка предистория.....	7
3. Логически модели на потвърдението	8
3.1. Хипотетико-дедуктивизъм	8
3.2. Потвърждение от отделни случаи	9
4. Вероятностни модели на потвърдението	10
4.1. Основни понятия в теория на вероятностите.....	11
4.2. Бейсианска теория на потвърдението (ВСТ)	13
5. Абдукция и извод в полза на най-доброто обяснение (IBE)	16
5.1. Концепцията за абдукция на Чарлс Пърс	16
5.2. Съвременното понятие за абдукция	17
5.3. Критики на абдукцията и аргументи в нейна защита	17
5.4. Абдуктивни модели.....	20
5.5. Видове обяснения	20
6. Обобщение: предимствата на бейсианската теория на потвърдението и на извода в полза на най-доброто обяснение пред другите подходи към потвърдението.....	23
III. Съвременната дискусия за връзката между бейсианската теория на потвърдението и извода в полза на най-доброто обяснение.....	24
1. Инкомпатибилисти	24
2. Компатибилисти.....	25
3. Бейсиански мерки за обяснителна сила.....	26
3.1. Емпирични изследвания	27
3.2. Критики.....	27
4. Обобщение: съвместимостта на бейсианската теория на потвърдението и извода в полза на най-доброто обяснение	28
IV. Ролята на приоритните вероятности в оценяването на обяснителната сила	29
1. Фиктивни примери	29
2. Примери от научната практика	32
2.1. Алфред Вегенер и теорията за континенталния дрейф	32

2.2. Батигин и Браун и хипотезата за „Планета девет“	36
3. Метод за изследване на условията на съвместимост между бейсианската теория на потвърждението и извода в полза на най-доброто обяснение.....	40
4. Обобщение: приорните вероятности и извода в полза на най-доброто обяснение.....	43
V. Заключение.....	44
Справка за приносите.....	46
Публикации по темата на дисертацията	47
Използвана литература.....	48

I. Увод

В уводната си част, дисертацията представя следния проблем: в съвременната философия на науката и по-точно в нейната поддисциплина – теория на потвърждението, паралелно съществуват два, на пръв поглед противоречиви, подхода към потвърждаването на научни хипотези и теории. По-популярният подход е т.нар. „бейсианска теория на потвърждението“ (*Bayesian Confirmation Theory* или ВСТ), която е изградена с формалния апарат на теория на вероятностите. Като алтернатива на бейсианската теория на потвърждението се явява теория, основана на т.нар. „извод в полза на най-доброто обяснение“ (*Inference to the Best Explanation* или просто ИВЕ). Основната разлика между двата подхода е в методиката – докато ВСТ разчита изцяло на емпирични свидетелства за потвърждаването на емпирични хипотези и теории, ИВЕ се основава на съображения относно качеството на обяснението (*explanatory considerations* или ЕС). С други думи, според ВСТ, трябва да бъде приписана по-висока степен на потвърденост на дадена хипотеза, тогава когато са налични емпирични свидетелства, *повишаващи нейната вероятност*. Според ИВЕ обаче, трябва да бъде приписана по-висока степен на потвърденост на дадена хипотеза (или пък дори хипотезата да бъде обявена за истинна), тогава когато тя *обяснява* наличните емпирични свидетелства по-добре от всички други конкуриращи се хипотези.

Основата на представения проблем е в това, че при наличието на два фундаментално различни подхода към потвърждението, естествено възниква въпросът за тяхната съвместимост. По този въпрос философите на науката са разделени, най-общо казано, на два лагера. Така наречените „инкомпатибилисти“ (*incompatibilists*) смятат, че двата подхода са несъвместими и още повече – че само единият от тях е рационален, докато другият трябва да отпадне. Противно на тях, компатибилистите (*compatibilists*) виждат различни възможности за съвместимост на двата подхода. Както инкомпатибилистите, така и компатибилистите обаче разглеждат въпроса за съвместимостта на ВСТ и ИВЕ като дихотомен – или двата подхода са напълно несъвместими, или те са напълно съвместими.

Дисертацията очертава възможността за трета позиция – условната съвместимост между двата подхода. По-конкретно, вниманието в дискусиата следва да се отмени от ниво „подход“ и да се фокусира на ниво отделен „модел“ на потвърждението. Тук могат да бъдат сравнени различни модели на ВСТ и различни модели на ИВЕ със задачата да се открие кои от тях са съвместими, в смисъла на това да дават еднакви резултати, когато бъдат приложени към един и същи случай, и *при какви условия* са съвместими те. В

дисертацията също така е предложен формален метод за определянето на въпросните условия на ВСТ-IBE съвместимост. Този метод взема два модела на потвърдението – един от ВСТ и един от IBE и ги прилага към конкретен случай, проверявайки какви условия трябва да бъдат изпълнени, за да дадат двата модела едни и същи резултати.

Успешното приложение на предлагания метод обаче е възможно само при една ясна формална експликация на основната идея зад IBE, а в съвременната литература върху IBE няма консенсус относно неговата експликация. Този проблем е обвързан с друг проблем пред IBE – то не разполага с формален механизъм за откриване на най-доброто обяснение. С други думи, IBE не дава информация за това как да бъде избрано най-доброто сред множество съревноваващи се обяснения, за да бъде направен извод в негова полза. Предложеното решение на тези проблеми в дисертацията е ключовото за IBE понятие „най-добро обяснение“ да бъде формално експлицирано, като основната интуиция зад тази негова експликация е, че „най-доброто обяснение сред всички съревноваващи се обяснения е онова, което има най-голяма *обяснителна сила*“. Обяснителната сила може да бъде квантифицирана чрез бейсианска мярка на обяснителната сила, която позволява и сравняване на обяснителната сила на различни хипотези. Ако E е адекватна мярка на обяснителната сила, „най-доброто обяснение“ ще бъде онова, което има най-висока стойност по E .

Дисертацията обаче показва, че съществуващите в литературата директни бейсиански мерки на обяснителната сила не са адекватни за вероятностната експликация на понятието „най-добро обяснение“, защото прилагането им в различни видове случаи води до незадоволителни резултати. Това е показано както за изкуствени примери, така и за случаи от реалната научна практика. В изкуствените примери, приложението на съществуващите мерки води до силно контраинтуитивни резултати, а в случаите от научната практика – до резултати, които се разминават с преценките на експертите. С оглед на този проблем е направено предложението вероятностната експликация на понятието „най-добро обяснение“ да отчита и приорните вероятности на оценяваните хипотези. Това предложение е подкрепено със същите примери, които демонстрират и слабостите на съществуващите мерки на обяснителната сила.

1. Централна теза

Докторската дисертация изгражда аргументи в полза на твърдението, че понятието „най-добро обяснение“, което е ключово за теорията на потвърдението, основаваща се на

„извод в полза на най-доброто обяснение“ (IBE), може да получи адекватна вероятностна експликация чрез бейсианска мярка на обяснителната сила, само ако последната отчита влиянието на приорните вероятности.

2. Основни цели на изследването

Дисертацията си поставя четири основни цели:

- i) Да представи ясно и подробно съвременния дебат във философията на науката относно съвместимостта на IBE и ВСТ;
- ii) Да покаже, че обяснителната сила и съответно понятието “най-добро обяснение”, които са ключови за IBE, подлежат на вероятностна експликация;
- iii) Да покаже предимствата на вероятностната експликация, отчитаща ролята на приорните вероятности, пред вече предлаганите формални експликации на обяснителната сила. Тези предимства ще бъдат демонстрирани чрез:
 - a. Фиктивни примери (*toy problems*);
 - b. Епизоди от история на науката (*case studies*);
- iv) Да въведе метод за изследването на условията за съвместимост между двете теории на потвърждението – IBE и ВСТ.

3. Метод на изследването

Дисертацията използва традиционната за философия на науката методология, която включва формалната експликация и изследването на случаи от история на науката. Формалната експликация, най-общо казано, се състои в превеждането на понятия или системи от понятия от една теоретична система на езика на по-богата теоретична система. Изследването на случаи от история на науката представлява детайлен анализ на определени епизоди от научната практика, с цел да се провери дали определен философски модел или концепция могат да отчетат и/или обяснят изследвания случай. В дисертацията има и части, които си служат със сравнителен анализ. Неговата роля е да сравнява по предварително зададени критерии два или повече модела на една и съща система, с цел да определи по-добрият от тях.

4. Структура на изложението

Дисертацията се състои от увод, три глави, заключение и кратко приложение.

Библиографията съдържа 139 заглавия на български и английски език.

II. Теория на потвърждението

1. Основи на теорията на потвърждението

Настоящата глава въвежда в теория на потвърждението по две направления – от една страна, с представянето на логическите и вероятностните модели на потвърждението, а от друга, с представянето на абдуктивните модели, сред които е и ИВЕ.

Теорията на потвърждението е клон от съвременната обща философия на науката, който изследва връзката на потвърждение между научните хипотези и теории и наличните за тях емпирични свидетелства. В теория на потвърждението има различни видове формални модели на потвърждението, които по математически и логически начин оказват как да определяме степента на потвърденост на определени хипотези и как да сравняваме съревноваващи се хипотези на базата на техните степени на потвърденост.

Основните компоненти на тези модели се обозначават по следния начин:

- хипотеза: h_i , където $h_i \in [h_1, h_n]$;
- емпирично свидетелство: e_i , където $e_i \in [e_1, e_n]$;
- познавателен фон: k_i , където $k_i \in [k_1, k_n]$;

където i обозначава конкретна хипотеза, емпирично свидетелство или познавателен фон, а n са общият брой на наличните хипотези, свидетелства и познавателни фонове.

Три вида отношения на потвърждение могат да бъдат обособени, в зависимост от вида на връзката между свидетелства и хипотеза : (i) *качествени* – „ e потвърждава h “; (ii) *количествени* – „ e потвърждава h в някаква степен“; (iii) *сравнителни* – „ e потвърждава h_1 в по-голяма степен, отколкото h_2 (h_3 , h_4 и т.н.)“ (Schurz, 2014). В зависимост от градацията на връзката между свидетелства и хипотеза пък, може да бъде направено разграничение между понятията „потвърждение“ и „верификация“, и „опровержение“ и „фалсификация“ (Nempel, 1945a; 1945b). „Верификация“ обозначава крайния случай на категорично потвърждение, а „фалсификация“ – крайния случай на категорично опровержение. Понятията „потвърждение“ и „опровержение“, от друга страна, допускат множество междинни степени на потвърденост.

2. Кратка предистория

Теорията на потвърждението възниква в средата на XX век, но философският проблем за потвърждаването на хипотетични твърдения се корени още в Античността. Той

занимава и Аристотел, който предлага дедуктивен модел на потвърдението – дадено хипотетично твърдение е потвърдено, ако то логически следва от определени необходимо истинни „първи принципи“. Този модел доминира Западното мислене до XVII век, когато е изместен от ранните индуктивни модели на потвърдението, според които не първите принципи, а опитът е в основата на потвърждаването на научните хипотези. Индуктивните модели обаче се сблъскват с проблема на Хюм за индукцията: от наблюдавани явления e_1, \dots, e_n , които имат някаква обща характеристика, не може със сигурност да бъде направен извод, запазвайки истинността, че такава характеристика ще има и още ненаблюдаваното, но сходно явление e_{n+1} . Казано по друг начин, индукцията, за разлика от дедукцията, не може да предостави *сигурно* знание.

Търсейки отговор на проблема на индукцията, в първите десетилетия на XX век моделите на потвърдението се разделят в две основни направления – логически модели и вероятностни модели. Към логическите модели спадат хипотетико-дедуктивното потвърждение, развито в детайли от Попър (Popper, 1935/2005) както и потвърдението от отделни примери, предлагано от Хемпел (Hempel, 1945a; 1945b). Към вероятностите модели спадат всички бейсиански модели на потвърдението.

3. Логически модели на потвърдението

3.1. Хипотетико-дедуктивизъм

Тази част от настоящата глава разглежда хипотетико-дедуктивният модел, който е представен за пръв път от Уилям Хюъл (Whewell, 1837), а най-известният му поддръжник е Попър (Popper, 1935/2005). В отговор на проблема на индукцията, хипотетико-дедуктивизмът на Попър отхвърля възможността за индуктивно потвърждение на научните хипотези и теории. Служейки си с термините „верификация“ и „фалсификация“, Попър настоява на това, че научните хипотези не могат да бъдат верифицирани, на базата на подкрепящи ги емпирични свидетелства – те могат да бъдат само фалсифицирани на базата на опровергаващи свидетелства. Основната задача на добрия учен, според това гледище, е да създава хипотези с високо емпирично съдържание, т.е. такива, от които могат да бъдат изведени много нетривиални прогнози, след което да тества тези хипотези емпирично, в опит да ги опровергае. Хипотезите, които временно устояват на тестовете, са „короборирани“ – те не са категорично потвърдени, но няма рационална причина да не бъдат използвани, докато не се появи по-добра алтернатива или те не бъдат фалсифицирани.

Съвременният хипотетико-дедуктивизъм (*hypothetico-deductivism* или HD) се разграничава от термините „верификация“, „фалсификация“ и „короборация“, като вместо тях си служи с „потвърждение“ и „опровержение“. Класическата форма на HD определя дали дадена хипотеза h е потвърдена от наличните свидетелства e , при някакъв познавателен фон k , по следния начин:

- Свидетелство e потвърждава хипотеза h при наличието на познавателен фон k , тогава и само тогава, когато (Schurz, 2014; Crupi, 2013):
 - От h и k следва e , т.е. $(h \wedge k) \models e$;
 - h е консистентно с k , т.е. $\{h, k\} \not\models \perp$;
 - e има съдържание, при дадено k , т.е. $k \not\models e$.
- e опровергава h при k , тогава и само тогава, когато:
 - От h и k следва $\neg e$, т.е. $(h \wedge k) \models \neg e$;
 - h е консистентно с k , т.е. $\{h, k\} \not\models \perp$;
 - $\neg e$ има съдържание, при дадено k , т.е. $k \not\models \neg e$.
- e е неутрално за h при k във всички останали случаи.

Класическата форма на HD се сблъсква с множество проблеми от теоретично естество, сред които са: тезата на Дюем-Куайн (Duhem, 1906/1954; Quine, 1951), “парадоксът на нерелевантната конюнкция” (Glymour, 1980), *reductio ad absurdum* за HD (Goodman, 1983) и “парадоксът на концептуалната разлика” (Laudan and Leplin, 1991, 464 стр.). В допълнение, HD не разполага с формални механизми за разграничаване между по-очаквани и по-слабо очаквани емпирични свидетелства, както и за различаването на все още неопровергани научни хипотези по отношение на тяхната потвърденост.

Съвкупността от тези проблеми пред HD води до относителното му изоставяне, в полза на бейсианските модели на потвърдението.

3.2. Потвърждение от отделни случаи

Тази част от настоящата глава разглежда Хемпеловото потвърждение от отделни случаи, което може да бъде експлицирано по следния начин: нека a и b са константи, съставляващи твърдението за емпиричните свидетелства e , т.е. a и b не са променливи, а обозначават конкретни факти, и e твърди за тях нещо нетривиално (не например $Q(a) \vee \neg Q(a)$). Нека множеството на елементите в твърдението за наличието на емпирични свидетелства e обозначим с $S(e)$. В нашия случай, $S(e) = \{a, b\}$. Онова, което хипотезата

казва именно за фактите a и b , т.е. “разгръщането” (*development*) на хипотезата h върху $S(e)$, ще бъде обозначено с $dev_e(h)$ (*e-development of h*). За една хипотеза от типа $\forall x(P(x) \rightarrow Q(x))$ и $S(e) = \{a, b\}$, $dev_e(h)$ ще бъде: $(P(a) \rightarrow Q(a)) \wedge (P(b) \rightarrow Q(b))$.

За всеки h , e и a , които принадлежат на множеството на изреченията от формален език от първи ред и e е логически непротиворечиво, съдържащо само константи (без променливи или квантор), е валидно следното (Crupi, 2013):

- e директно Хемпел-потвърждава h , тогава и само тогава, когато $e \models dev_e(h)$;
 - e индиректно Хемпел-потвърждава h , тогава и само тогава, когато $e \models dev_e(a)$ и $a \models h$;
- e директно Хемпел-опровергава h , тогава и само тогава, когато $e \models dev_e(\neg h)$;
 - e индиректно Хемпел-опровергава h , тогава и само тогава, когато $e \models dev_e(a)$ и $a \models \neg h$;
- e е Хемпел-неутрално за h във всички други случаи.

В този си вид, потвърждението от отделни случаи се сблъсква с поне два сериозни проблема: Хемпеловият „парадокс на гарваните“ и т.нар. „нова загадка на индукцията“, известна още като „парадокс на зелиното“ (*the grue paradox*) на Нелсън Гудман (Goodman, 1983). Опитвайки да се справи с тези проблеми, Хемпел първоначално въвежда в своя модел формален механизъм за определяне на *степени* на потвърждение (Hempel and Oppenheim, 1945). В крайна сметка обаче, Хемпеловият логически модел на потвърждението също бива изоставен в полза на вероятностните модели, които могат да боравят със степени на потвърждението естествено и с лекота.

4. Вероятностни модели на потвърждението

Дисертацията продължава с въведение в основите на теория на потвърждението и представянето на вероятностните модели на потвърждението. В съвременната теория на потвърждението, вероятностните модели са по-известни като “бейсиански модели на потвърждението” (*Bayesian models of confirmation*), което означава че: i) те използват експлицитно или имплицитно теоремата на Бейс; ii) те възприемат епистемическата интерпретация на теория на вероятностите.

4.1. Основни понятия в теория на вероятностите

Съвременната класическа теория на вероятностите получава своята аксиоматизация от руския математик Андрей Колмогоров (Kolmogorov, 1933/1956). Нейните аксиоми са:

- (1) Неотрицателност (*non-negativity*): $P(A) \geq 0$;
- (2) Нормираност (*normalization*): $P(\Omega) = 1$;
- (3) Адитивност (*finite additivity*): $P(A \vee B) = P(A) + P(B)$, за всеки A и B , такива че $(A \wedge B) = \emptyset$.

На естествен език, аксиомите гласят:

- (1) Вероятността на което и да е събитие A има за стойности естествени числа по-големи или равни на 0.
- (2) Вероятността на множеството от всички възможни резултати е 1.
- (3) За всеки две несъвместими събития A и B , вероятността да се реализира A или B е равна на сборът на вероятностите на A и B .

Вероятността $P(-)$ е *функция*, която определя числови стойности на някакви *събития*. “Събитие” в теория на вероятностите е множество от *резултати* от даден *опит*. “Опит” пък е всяка процедура с добре дефинирано множество от възможни резултати. Опитите с един единствен възможен резултат се наричат “детерминистични” (*deterministic*). Опитите с два или повече възможни резултата са “случайни” (*random*). Множество, съдържащо само един резултат, се нарича “елементарно събитие” (*elementary event*). Множествата, съдържащи повече от един резултат са “сложни събития” (*complex events*). Празното множество, т.е. множеството, което не съдържа нито един резултат, също е събитие (с нулева вероятност). Събитията, на свой ред, са подмножества на *пространството на елементарните събития* (Ω).

“Несъвместими” се наричат онези събития, които не могат да се реализират едновременно. Събитията, които могат да се реализират едновременно, се наричат съответно “съвместими”. Правилото за събиране на вероятности за всеки две съвместими събития A и B има следната формула:

$$P(A \vee B) = P(A) + P(B) - P(AB)$$

Две събития са “независими” (още “вероятностно независими”), тогава и само тогава, когато реализирането на едното не променя вероятността за реализирането на другото. Правилото за умножаване вероятностите на *зависими* събития, има следната формула:

$$P(AB) = P(A) \times P(B|A)$$

Вероятността $P(B|A)$ се нарича “условна вероятност” и гласи: “вероятността на В при дадено А”.

“Противоположни събития” се наричат събитията, при реализирането на едното от които, другото не може да се реализира, и обратното. Вероятностите на противоположните събития винаги имат сбор 1. Тук в дисертацията е отбелязано и това, че *несъвместимите* събития не са задължително *противоположни*, както и че $P(A|B)$ – вероятността на А при дадено В, не е задължително равна на $P(B|A)$ – вероятността на В при дадено А.

Върху математическия аспект на теория на вероятностите, изложен в дисертацията, има относителен консенсус, но такъв липсва по въпроса за нейната философска интерпретация. Съществуват няколко различни философски интерпретации на теория на вероятностите, които могат да бъдат разделени на два основни вида: обективистка (в някои среди известна като “физическа” или “честотна”) и субективистка (още “епистемическа”). Към обективистките интерпретации спадат (Gillies, 2000):

- a) Класическата интерпретация – вероятността е отношението между резултатите, които представляват интерес, и общия брой възможни равновероятни резултати от даден опит. Вероятността на събитието А – да бъде изтеглено асо ($m = 4$) от добре разбъркано пълно тесте карти ($n = 52$), е:

$$P(A) = m/n = 4/52 \approx 0.077$$

- b) Логическата интерпретация – вероятността е вид логическо отношение, тя изразява степента на недедуктивно (т.е. частично) следствие на някакви твърдения от други твърдения. Ако „всеки гарван, който е наблюдаван, е бил черен“, то от истинността на това твърдение следва *до някаква степен* истинността на твърдението „всички гарвани са черни“.
- c) Честотната интерпретация – вероятността изразява относителната честота, с която някакво събитие се проявява в безкрайна поредица от опити. При хвърлянето на добре балансирана монета безброй пъти, относителната честота на ези ще бъде 50%.

За разлика от обективистките интерпретации, субективистката интерпретация възприема вероятността като характеристика на *собствените убеждения* относно някакви събития,

а не като характеристика на самите събития. Към субективистката интерпретация спадат множеството съвременни разновидности на бейсианството.

4.2. Бейсианска теория на потвърждението (ВСТ)

Формалният механизъм на математическата теория на вероятностите представлява основата на всички бейсиански модели на потвърждението. В дисертацията вече бе отбелязано, че тези модели използват имплицитно или експлицитно теоремата на Бейс. За всяка двойка събития A и B , такива, че $P(B) \neq 0$ теоремата на Бейс има следната формула:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

На естествен език: вероятността на A при дадено B е равна на произведението на условната вероятност на B при дадено A и вероятността на A , разделено на вероятността на B .

За дадена хипотеза h и емпирични свидетелства e , теоремата на Бейс гласи:

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$

На естествен език: вероятността хипотезата h да е истинна, при наличието на емпирични свидетелства e , може да бъде установена като се намери произведението на вероятността да са налични такива емпирични свидетелства, ако хипотезата е истинна, и вероятността ни хипотезата да е истинна по начало (преди свидетелствата да са налични), разделено на вероятността такива свидетелства да са налични въобще.

Отделните компоненти в горната формула имат своите технически определения. Те са, както следва:

- $P(h)$ – “приорна вероятност” (*prior probability*) или вероятността, която бива приписана на хипотезата h , преди да бъде установено наличието на свидетелствата e .
- $P(h|e)$ – “постериорна вероятност” (*posterior probability*) или вероятността, която бива приписвана на хипотезата h , след като бъде взето предвид наличието на свидетелствата e .
- $P(e|h)$ – “очакваност на свидетелствата” (*likelihood*) или вероятността да са налични свидетелствата e , ако хипотезата h е истинна.

- $P(e)$ – вероятност на свидетелствата, изразяваща доколко вероятно е наличието на свидетелствата e , когато бъдат взети предвид всички възможни хипотези относно тях.

Теоремата на Бейс не може да даде информация за това колко вероятна е определена хипотеза първоначално – тази информация следва да бъде въведена във формулата. Бейсианците са разделени по въпроса как следва да бъде добита подобна информация: според бейсианците обективисти, има определени логически правила, които трябва да бъдат следвани при определянето на стойности на приорните вероятности, докато според бейсианците субективисти, приписването на каквито и да било приорни вероятности е позволено, стига те да спазват аксиомите на теория на вероятностите.

В дисертацията са разгледани както качествени, така и количествени бейсиански модели на потвърдението. Примери за качествени модели са (Cupri, 2013):

Потвърждение от твърдост (*firmness*) на убежденията (*F*-потвърждение)

- e *F*-потвърждава h , спрямо k , тогава и само тогава, когато $P(h|e \wedge k) > 0.5$;
- e *F*-опровергава h , спрямо k , тогава и само тогава, когато $P(h|e \wedge k) < 0.5$;
- e е *F*-неутрално за h , спрямо k , тогава и само тогава, когато $P(h|e \wedge k) = 0.5$.

На естествен език, хипотезата h е *F*-потвърдена от емпиричните свидетелства e , при наличието на познавателен фон k , когато нейната вероятност надхвърля някакъв вероятностен праг (в случая 0.5); хипотезата е *F*-опровергана от свидетелствата, ако вероятността ѝ е под този праг; а в случаите, в които вероятността на хипотезата е точно на прага (т.е. = 0.5), свидетелствата не са релевантни за хипотезата.

Потвърждение от релевантност (*relevance*) на свидетелствата (*R*-потвърждение)

- e *R*-потвърждава h , спрямо k , тогава и само тогава, когато $P(h|e \wedge k) > P(h|k)$;
- e *R*-опровергава h , спрямо k , тогава и само тогава, когато $P(h|e \wedge k) < P(h|k)$;
- e е *R*-неутрално за h , спрямо k , тогава и само тогава, когато $P(h|e \wedge k) = P(h|k)$.

Казано иначе, хипотезата h е *R*-потвърдена от емпиричните свидетелства e , при наличието на познавателен фон k , когато нейната постериорна вероятност е по-висока от нейната приорна вероятност; хипотезата е *R*-опровергана от свидетелствата, ако постериорната ѝ вероятност е по-ниска от приорната; а в случаи, в които постериорната и приорната вероятност на хипотезата са равни, свидетелствата не са релевантни за хипотезата (т.е. те изобщо не променят нейната начална вероятност).

Количествените байсиански модели са под формата на вероятностни мерки на потвърждението. Общото за всички тези мерки е, че потвърждението се явява някакъв вид вероятностна функция, нарастваща с наличието на подкрепящи хипотезата свидетелства. Основните байсиански мерки на потвърждението са представени в Таблица 1 (Tentori et al., 2007).

Таблица 1. Байсиански мерки на потвърждение

Мярка на потвърждение	Източник
$C_1(e, h) = P(h e) - P(h)$	Eells, 1982; Jeffrey, 1992
$C_2(e, h) = P(e h) - P(e \neg h)$	Nozick, 1981
$C_3(e, h) = \log \left[\frac{P(h e)}{P(h)} \right]$	Keynes, 1921/2004; Horwich, 1982
$C_4(e, h) = \log \left[\frac{P(e h)}{P(e \neg h)} \right]$	Good, 1984
$C_5(e, h) = P(h \wedge e) - [P(e) \times P(h)]$	Carnap, 1950/1962
$C_6(e, h) = \frac{P(e h) - P(e \neg h)}{P(e h) + P(e \neg h)}$	Kemeny and Oppenheim, 1952
$C_7(e, h) = \begin{cases} \frac{P(h e) - P(h)}{1 - P(h)} & \text{ако } P(h e) \geq P(h) \\ \frac{P(h e) - P(h)}{P(h)} & \text{ако } P(h e) < P(h) \end{cases}$	Crupi, Tentori and Gonzalez, 2007; Crupi and Tentori, 2013

Бейсианската теория на потвърждението също има своите проблеми от теоретично естество, сред които са: проблемът от несигурност на свидетелствата, проблемът от старите свидетелства, проблемът за определянето на приорите и предпоставката, че когнитивните агенти са логически всезнаещи (Talbot, 2001). Тези проблеми допринасят за разнообразието от байсиански модели на потвърждението, като първият от тях мотивира и създаването на следната генерализация на теоремата на Бейс, известна като „кондиционализация на Джефри“ (*Jeffrey conditionalization* или JC) (Jeffrey, 1990):

$$JC: P_f(h) = P_i(h|e) \times P_f(e) + P_i(h|\neg e) \times P_f(\neg e)$$

В горната формула f (от *final*) обозначава *финална вероятност*, т.е. $P_f(h)$ е вероятността на хипотезата h , при налични свидетелства e , а $P_f(e)$ е финалната вероятност на свидетелствата при извършено наблюдение и установяване на тяхната наличност. Също така, i (от *initial*) обозначава *начална вероятност*: $P_i(h)$ е първоначалната вероятност на

хипотезата, а $P_i(e)$ е вероятността свидетелствата да са налични, преди все още да е извършено наблюдение, установяващо наличността им.

Дисертацията показва, че за случаи, в които $P_f(e) = 1$ (следователно $P_f(\neg e) = 1 - P_f(e) = 0$), JC се опростява до правилото на Бейс. В този смисъл, правилото на Бейс е частен случай на JC, в който за финална вероятност на свидетелствата е приписана крайната стойност 1. Основното предимство на JC е именно в това, че то се справя в случаи, в които няма пълна сигурност в свидетелствата ($P_f(e) < 1$), т.е. JC преодолява проблема от несигурност на свидетелствата (за разлика от самото правило на Бейс). В съвременната теория на потвърдението, JC е най-популярното вероятностно правило за рационална ревизия на убежденията.

5. Абдукция и извод в полза на най-доброто обяснение (IBE)

5.1. Концепцията за абдукция на Чарлс Пърс

Следващата част от настоящата глава разглежда понятието за абдукция и критиките към него, както и различните видове абдуктивни модели и основните концепции за обяснението, които залягат в тях. Изложението започва с Чарлс Пърс, който въвежда понятието „абдукция“, с цел да реферира единствения според него метод за откриването на нови научни хипотези и теории. За разлика от съвременното понятие „абдукция“, Пърсовата „абдукция“ принадлежи изцяло на контекста на откритието и няма никакви потвърждаващи хипотезата функции. Според това гледище, абдукцията е вид еволюирал човешки инстинкт, който открива истинното обяснение сред огромно количество възможни обяснения на даден феномен. Като вид човешки инстинкт, Пърсовата абдукция не се нуждае от нормативна обосновка. Въпреки това, тя има логическа структура, според която, ако дадено обяснение прави конкретен наблюдаван феномен *по-малко изненадващ* (т.е. *по-очакван*), то имаме причина да смятаме, че обяснението е истинно.

Пърсовото понятие за абдукция среща критика поне по две основни направления. Едното от тях е, че понятието не е ясно експлицирано. Във варианта, представен от Пърс, то се оказва по-скоро неясна обща интуиция, отколкото точно правило за правенето на рационален извод. Вторият вид критика е, че ролята на абдукцията като метод за правенето на рационални изводи не е безпроблемно определена. Ако нейната роля е да открива нови хипотези, то тя не може задоволително да изпълнява тази роля, защото тя

разчита на вече *налична* хипотеза, която да прави някакъв наблюдаван феномен по-очакван. Ако пък ролята на абдукцията е да бъде филтър, чрез който биват отсеци онези хипотези, които правят някакъв феномен по-очакван, с намерението те да бъдат подложени на емпиричен тест, то проблемът е, че в повечето случаи има безкрайно много такива хипотези. Ролята на абдукцията като филтър за хипотези в такива случаи изглежда се обезсмисля.

Макар че както ялната експликация, така и ролята на абдукцията да са обекти на спор, дисертацията показва, че абдукцията е широко разпространена, както в ежедневието, така и в науката. Тя бива използвана в широк спектър от човешки дейности, простиращ се от компютърните науки и изкуствения интелект, през формирането на исторически хипотези, до изграждането на различни философски аргументи.

5.2. Съвременното понятие за абдукция

Абдукцията, също както логическата индукция, е немонотонен амплиативен метод за правенето на рационални изводи. С други думи, абдукцията позволява ревизия на истинността на заключенията си, с прибавянето на нови предпоставки (немонотонност), а също така позволява нейният извод да предоставя по-широко концептуално познание от информацията, съдържаща се в предпоставките ѝ (амплиативност). Ключова разлика между двата метода обаче е това, че абдукцията експлицитно или имплицитно се осланя на *съображения относно качеството на обяснението* (ЕС), а индукцията се осланя основно на *наблюдавани честоти на проява на някакви събития*. Класическата и относително наивна дефиниция на абдукцията гласи следното:

A1: „При наличието на свидетелство *e* и възможни обяснения [*candidate explanations*] h_1, \dots, h_n , да се направи извод в полза на истинността на онази h_i , която най-добре обяснява *e*.“ (Douven, 2011)

5.3. Критики на абдукцията и аргументи в нейна защита

Бас ван Фраасен (Van Fraassen, 1989) е един от най-известните критици на абдукцията, като неговите критики биват обобщени в три основни аргумента. Първият гласи, че ако абдукцията бъде приета за „извод в полза на най-доброто обяснение“ (IBE), то нищо не гарантира наличието на „най-добро обяснение“ в множеството на възможни обяснения, които са достъпни в даден момент. Съвсем възможно е, и доста вероятно, „най-доброто обяснение“ изобщо да не е налице и в този смисъл да е извън множеството на

разглежданите в даден момент обяснения. Всъщност, възможно е въпросното множество да включва само лоши или дори *изключително* лоши обяснения на наличните свидетелства и най-доброто измежду тези обяснения да е, обективно погледнато, също доста лошо.

Вторият аргумент на Ван Фраасен срещу IBE почива върху следната предпоставка: ако една хипотеза е *по-обяснителна* от друга, то тя трябва да е и *по-информативна* от другата. По-информативната хипотеза обаче не може да бъде *по-вероятно истинна* от по-малко информативната. Следователно, по-обяснителната хипотеза не може да бъде по-вероятна или по-добре потвърдена от по-малко обяснителната (противно на IBE).

Третият аргумент на Ван Фраасен срещу IBE пряко засяга съвместимостта му с бейсианската теория на потвърдението. Той обхваща само вероятностната версия на IBE и гласи следното: всеки модел на вероятностна абдукция трябва или да се подчинява на правилото на Бейс и в този смисъл да бъде редундантен (т.е. излишен), или да се отклонява от правилото на Бейс и в този смисъл да се изправи срещу т.нар. „аргумент от холандски залог“ (*Dutch book argument*). Казано накратко, аргументът от холандски залог показва, че всеки, който не следва аксиомите на класическата теория на вероятностите, не е рационален. В този смисъл Ван Фраасен твърди, че една вероятностна версия на IBE трябва да се подчинява на правилото на Бейс и да следва аксиомите на математическата теория на вероятностите, в който случай IBE става излишно, или пък то трябва да погазва въпросните аксиоми, в който случай използващите IBE са просто ирационални.

Самият Ван Фраасен допуска срещу своите критици на IBE три възможни вида отговор, изразяващи се в позициите, които той нарича „привилегия“, „форсмажор“ и „окопаване“ (*privilege, force majeure, retrenchment*).

- А) Позицията, наречена „привилегия“ (*privilege*) се изразява в предположението, че хората по природа имат способността да улучват такива множества от хипотези, които съдържат в себе си истинната хипотеза.
- В) „Форсмажор“ (*force majeure*) е позицията, защитаваща твърдението, че по принуда трябва да бъде избрано обяснение измежду наличните в исторически план хипотези, дори и те да представляват множество от слаби обяснения. От едно добро правило за разсъждение не може да се очаква друго, освен да избира *най-добрата* от тези налични хипотези (именно това което прави IBE).

С) „Окопаване“ (*retrenchment*) е третата и последна позиция, която Ван Фраасен допуска в защита на ИВЕ. Тя се състои в твърдението, че словосъчетанието „извод в полза на най-доброто обяснение“ (“*inference to the best explanation*”) е подвеждащо и не описва добре абдуктивното правило, към което се отнася.

Ван Фраасен в крайна сметка отхвърля и трите позиции, на базата на аргументи, посочващи тяхната незадоволителност.

Дисертацията разглежда и аргументи в защита на абдукцията срещу критиките на Ван Фраасен. Тук е разгледана дискусията около аргументите на Ричард Бойд (Boyd, 1980; 1983; 1985), които той предлага в защита на абдукцията и които са разновидност на т.нар. „аргумент от липсата на чудеса“ (*no-miracles argument*). Бойд твърди, че научните хипотези и теории стават все по-точни през годините, разчитайки на спомагателни научни теории. Последните обаче до голяма степен са формирани с помощта на абдукция. Следователно, има добри основания да се твърди, че абдукцията е изпитано и благонадеждно правило за правене на логически изводи. Връзката с аргумента от липсата на чудеса тук е, че ако абдукцията не бе благонадеждно правило, то успехът на научните хипотези и теории би бил мистериозен и необясним, т.е. би бил нещо като чудо.

В отговор на аргументите на Ван Фраасен срещу абдукцията, Питър Липтън пък предлага (Lipton, 1991/2004) възможността да бъде избрано такова множество от хипотези, чийто членове обемат цялото логическо пространство. Това лесно може да бъде направено чрез включването на “всеобхватна” хипотеза (*catch-all hypothesis*), която представлява отрицание на всички вече известни хипотези по даден въпрос. Ако множеството от налични в даден момент хипотези бъде дефинирано като $\{h_1, \dots, h_n\}$, то дефинирана може да бъде и една всеобхватна хипотеза h_{n+1} със съдържание $\neg\{h_1 \vee \dots \vee h_n\}$, която да бъде включена към въпросното множество с резултат: $\{h_1, \dots, h_n, h_{n+1}\}$. Последното множество изчерпва цялото логическо пространство от възможни хипотези, или казано иначе, един от неговите членове *трябва* да бъде истинна хипотеза. Проблемът в случая е, че ако истинната хипотеза е всеобхватната h_{n+1} , то не е ясно какви са нейните емпирични следствия (ако тя въобще има такива), а обяснителната ѝ сила е твърде ниска, дори нищожна. Тази хипотеза е еквивалентът на това да се твърди, че нито една от всички вече известни хипотези не е истинна.

5.4. Абдуктивни модели

Тази част показва, че „изводът в полза на най-доброто обяснение“ (IBE) е само едно правило от множеството на абдуктивните правила, които намират оптимално приложение в различни случаи и имат различни проблеми и решения на тези проблеми. Сред множеството на абдуктивните правила са абдукцията на факт, абдукцията на закон и абдукцията на теоретичен модел, както и екзистенциалните абдукции от втори ред (абдукция на микроскопична част, аналогична абдукция, абдукция на хипотетична причина) и абдукциите на обща причина.

5.5. Видове обяснения

Тук е разгледан вторият от двата основни елемента на IBE, които са: някакъв вид (логически) извод (*inference*) и определено обяснение (*explanation*). Обяснението може да играе две различни роли в абдуктивния метод. Първата и очевидна роля е тази на *най-доброто* обяснение (*the best explanation*), като IBE представлява извод в негова полза. Втората роля, която обяснението може да играе като елемент на абдукцията, е тази на *потенциално* обяснение (*candidate explanation*). Потенциални обяснения са всички членове на множеството от обяснения на даден феномен, между които е и най-доброто обяснение.

В съвременната философия на науката доминира плурализъм относно въпроса: „кой е най-добрият модел на научно обяснение?“. С други думи, много съвременни философи на науката са по-скоро склонни да смятат, че няма един единствен модел на научно обяснение, който да е адекватен на всички случаи в науката, а има по-скоро множество добри модели на научно обяснение, които са приложими в конкретни случаи и зависят от контекста на приложение.

Дисертацията разглежда анализа на един от най-известните изследователи на обяснението – Питър Липтън (Lipton, 2001a), който обстойно сравнява различните модели на обяснение по техните силни и слаби черти. Той изхожда от предпоставката, че основната ценност на обяснението е това, че то допринася за „разбирането“ (*understanding*) на експланандума. Оттук той сравнява различните модели на обяснение (които приемат различни концепции за „разбиране“) по техните основни характеристики, като формира три критерия на сравнение. От един добър модел на обяснението се иска да може да (Lipton, 2001a, стр.43):

- a) прави разлика между това да *знаем* за появата на един феномен и да *разбираме* защо този феномен се проявява;
- b) дава възможност да предлагаме обяснения, които не са на свой ред обяснени;
- c) дава възможност да обясним даден феномен тогава, когато самият феномен играе ключова роля в нашето убеждение, че обяснението е правилно.

С други думи, първият критерий проверява дали едно обяснение може да запази разликата между „знание“ и „разбиране“. Вторият критерий проверява податливостта на въпросното обяснение към един безкраен регрес в търсенето на още и още обяснения (т.нар. „*why regress*“). Третият критерий проверява дали обяснението има обяснителна стойност, когато само предоставя свидетелства за себе си. Подобен вид обяснения са наречени от Хемпел „себезасвидетелстващи обяснения“ (*self-evidencing explanations*) (Hempel, 1965, стр. 372).

Липтън сравнява следните пет концепции на „разбиране“, формиращи различни видове обяснения (Lipton, 2001a, стр. 45-48):

- A) „Разбиране“ в смисъла на *предоставяне на добро основание за убеденост* („*reason*” *conception of understanding*). Ако експланандумът в случая е *P*, въпросът, на който този вид обяснение търси отговор е „Защо да вярвам, че *P*?“. Това определение на „разбиране“ избягва неясни метафизически концепции, разчитайки предимно на епистемически ключовото понятие „основание за убеденост“ (*reason for belief*).
- B) „Разбиране“ в смисъла на *свеждане до добре познатото* („*familiarity*” *conception of understanding*). Според тази концепция, основната функция на процеса на разбиране е да прави непознатото за нас – добре познато (*unfamiliar – familiar*), като често това става чрез свързването на непознатите и изненадващи факти с вече добре познати такива.
- C) „Разбиране“ в смисъла на *унификация* („*unification*” *conception of understanding*). Според тази концепция, да разберем *P* означава да видим как *P* се вписва в едно общо цяло с други познати за нас факти.
- D) „Разбиране“ в смисъла на *необходимост* („*necessity*” *conception of understanding*). Обясненията, които допринасят за нашето разбиране, според тази концепция, всъщност показват, че експланандумът е *трябвало* да бъде налице. С други думи, те обясняват, показвайки *необходимостта* от появата на обяснявания феномен.

Е) „Разбиране“ в смисъла на *каузалност* (“*causal*” *conception of understanding*).

Според тази концепция, да разберем *P* означава да дадем информация за причините за *P*.

В следствие на своя анализ Липтън заключава, че само каузалните обяснения, т.е. тези, които предоставят „разбиране“ в смисъла на *каузалност*, отговарят и на трите критерия за добър модел на обяснението.

Независимо от модела, към който дадени научни обяснения принадлежат (дали те са унификационистки, каузални и т.н.), при прилагането на ИВЕ трябва да бъде направен извод в полза на „най-доброто“ от тях. В този смисъл, за целите на дисертацията бива приет вид плурализъм относно *моделите* на обяснението, като дисертацията се фокусира върху въпроса как могат да бъдат сравнявани *отделни* обяснения.

Липтън (Lipton, 1991/2004; 2001a) предлага следното решение на този проблем. Според него, множеството от потенциални обяснения може да бъде (и в почти всички случаи е) на практика безкрайно, включващо изключително много напълно незадоволителни обяснения. Затова той предлага абдуктивен механизъм с два филтъра, чиято цел е да отсее най-доброто от всички налични обяснения. Първият филтър определя кое от наличните обяснения е *най-„хубавото“* (*the loveliest*), а вторият взема обясненията, които са преминали през първия филтър и определя кое от тях е *най-вероятното* (*the likeliest*). Механизмът наподобява старата схема на Пърс, в която абдукцията играе основна роля в изнамирането на онази хипотеза, която изглежда най-обещаваща, с цел бъде подложена на емпирична проверка.

Тук се поражда въпросът как може да бъде определено дали една обяснителна хипотеза е „хубава“, в смисъла на Липтън? Дисертацията излага най-разпространеното в литературата мнение, според което най-добрата обяснителна хипотеза (или най-„хубавото“ обяснение, по термините на Липтън) е онази, която се справя по-добре от всички останали обяснителни хипотези по няколко отличителни критерия. Сред тези критерии, най-често давани са следните: простота на обяснението (*simplicity*), унификация на обясняваните феномени (*unification*), обхват на обяснението (*scope*); като понякога тук влиза и кохерентност на убежденията относно обясняваните феномени (*coherence*). С други думи, идеалната обяснителна хипотеза е относително проста (например има малко на брой предпоставки или допускания), обяснява голямо количество феномени, като обединява много от тях в своето общо обяснение, и в

допълнение премахва несъвместимостта между различните убеждения относно обясняваните феномени, засилвайки по този начин кохерентността на убежденията.

Дисертацията разглежда и трудностите, които тези критерии за добро обяснение срещат. На първо място, някои от тях често са в обратнопропорционална зависимост (като критериите „простота“ и „обхват“ на обяснението, например). На второ място, самите критерии не са особено ясни и върху техните значения няма консенсус. На трето място, общият брой на тези критерии не е ясно установен – различните автори боравят с различен брой критерии.

6. Обобщение: предимствата на бейсианската теория на потвърдението и на извода в полза на най-доброто обяснение пред другите подходи към потвърдението

Дисертацията достига до заключението, че бейсианският подход е най-популярен в съвременната теория на потвърдението за момента, като вероятно тази тенденция ще се запази в близкото бъдеще. Може би една от най-важните причини за това е относителният провал на логическия подход към потвърдението. Основното предимство на бейсианската теория е, че нейните вероятностните модели решават проблемите пред логическите модели на потвърдението с относителна лекота, най-вече защото бейсианските модели естествено боравят с концепцията за *степени на потвърждение*, докато за логическия подход тази концепция е чужда и трябва да бъде привнесена отвън.

От друга страна обаче, множество философи и учени от Пърс насам имат силната интуиция, че потвърдението на дадена хипотеза/теория зависи не само от наличните *свидетелства*, но и от това колко добре тя *обяснява* тези свидетелства. Оттук и идеята за модел, който използва *обяснителната сила* на хипотезата, за да направи извод относно степента ѝ на *потвърждение*. Такъв пример е „извод в полза на най-доброто обяснение“ (IBE), който е най-популярният от множеството съществуващи видове абдуктивни модели.

Наличието на тези два различни подхода към потвърдението (ВСТ и IBE) естествено поражда въпроса за връзката между тях. Този въпрос е обект на интензивни изследвания и дебат в съвременната теория на потвърдението.

III. Съвременната дискусия за връзката между бейсианската теория на потвърдението и извода в полза на най-доброто обяснение

Тази глава от дисертацията разглежда дебатът относно съвместимостта между ВСТ и ИВЕ, а също и наличните до момента бейсиански мерки на обяснителната сила. На основата на заключенията от тази глава, в следващата е направено предложение за нов подход към въпроса за ВСТ-ИВЕ съвместимостта. На първо място е показано, че този въпрос разделя философите на науката, занимаващи се с потвърдението, на два лагера. От едната страна са т.нар. „инкомпатибилисти“, според които ВСТ и ИВЕ нямат бъдеще заедно. Инкомпатибилистите смятат, че двата подхода към потвърдението не могат да бъдат едновременно подкрепяни и обикновено отхвърлят рационалността на единия в полза на другия. От другата страна са т.нар. „компатибилисти“, според които двата подхода са частично или напълно съвместими. Компатибилистите са движени от мотивацията да запълнят методологичните пропуски в двата подхода, като ги обединят един с друг, показвайки, че те не се различават в своите резултати.

1. Инкомпатибилисти

Един от известните инкомпатибилисти е Бас ван Фраасен, чиято основна критика срещу съвместимостта между бейсианската теория на потвърдението и ИВЕ е неговият трети аргумент срещу ИВЕ (Van Fraassen, 1989). Аргументът, както бе представен в предната глава на дисертацията, е следният: използването на ИВЕ означава употреба на различно правило за ревизия на убежденията от правилото на Бейс. Това на свой ред означава, че или ИВЕ дава същите резултати, каквито дава правилото на Бейс, в който случай ИВЕ става тривиално и излишно, или ИВЕ дава резултати различни от правилото на Бейс, в който случай ИВЕ се сблъсква с т.нар. „аргумент от холандски залог“ (описан в предната глава).

Дисертацията разглежда и множеството аргументи за инкомпатибилизма на двата подхода на потвърдението, предложени от Уесли Салмън. Тук те ще бъдат изредени съвсем накратко. Първият аргумент е простото твърдение, че съвсем не е ясно какво означава „добро обяснение“ (Salmon, 2001a). Вторият гласи, че ИВЕ размива разликата между въпросите „защо“, търсеци *обяснение* и въпросите „защо“, търсеци *потвърждение*. Трета точка в критиката на Салмън е, че, за определянето на дадена обяснителна хипотеза като „добра“, едно от класическите изисквания е тя да бъде *истинна*, но това не може да бъде решено от ИВЕ, тъй като то разчита на вече *налична*

добра обяснителна хипотеза, за да направи своя извод в нейна полза. Обобщавайки своите аргументи, Салмън заключава, че *потвърждението е логически независимо от обяснението* (Salmon, 2001b).

2. Компатибилисти

Настоящата глава разглежда значително по-многобройните предложения на компатибилистите, които почиват на две основни идеи. Едната е, че съображенията относно качеството на обяснението (ЕС), които формират ИВЕ, биха могли да служат за определянето на приорната вероятност и/или т.нар. „очакваност на свидетелствата“ (*likelihood*) във формулата на Бейс. Тъй като ВСТ няма формален механизъм, по който да определя тези стойности, ИВЕ би могло да послужи като един вид допълнение на бейсианския механизъм. Това предложение приписва второстепенна роля на ИВЕ, оставяйки основната работа по потвърждението да се върши от бейсианската теория на потвърждението. От друга страна, предложението напълно интегрира двата подхода в един. Тук спадат предложенията на: Lipton, 1991/2004; 2001b; Okasha, 2000; Henderson, 2014 и др.

Втората основна идея на компатибилистите се поражда от това, че някои привърженици на ИВЕ (например Psillos, 2007) не могат да се примирят с отреждането на второстепенна роля за ИВЕ в теория на потвърждението. Според тях, ИВЕ е напълно равноправно с ВСТ и ролята, която то трябва да получи в теория на потвърждението, трябва да бъде съответстваща на този му ранг. Предложенията тук са две: или ИВЕ ще формира приорните вероятности в бейсианския механизъм, но *по нормативен начин* (с други думи, бейсианците субективисти ще трябва да станат обективисти), или ИВЕ ще остане независим метод и неговата съвместимост с ВСТ ще трябва да бъде показана по някакъв друг начин. Подобни предложения в най-добрия случай предлагат само частична интеграция между двата подхода.

В обобщение на позициите, които дисертацията разглежда по дебата за съвместимостта между ВСТ и ИВЕ, в нея е достигнато до заключението, че съществуващите за момента опити за *интеграция* на двата подхода към потвърждението се провалят. Нито едно от разгледаните предложения не е достатъчно убедително или пък ясно, за да реши проблема със съвместимостта на ВСТ и ИВЕ категорично в полза на инкомпатибилизма или компатибилизма. Затова в дисертацията е направено предложението да се отмести фокусът на проблема от цялостна *интеграция* на ВСТ и ИВЕ и да се премине към

разглеждане на възможността за формална *експликация* на ключовото за ИВЕ понятие “най-добро обяснение” чрез бейсианска мярка на обяснителната сила. Следвайки логиката на изложението, в следващата част са представени съществуващите директни бейсиански мерки на обяснителната сила.

3. Бейсиански мерки за обяснителна сила

Основната идея зад различните опити за експликация на понятието „обяснителна сила“ е да направят възможно измерването на това колко обяснителна сила притежава дадена хипотеза или с други думи, доколко добре дадена хипотеза обяснява наличните емпирични свидетелства. Интуицията зад тези опити е, че *обяснителната сила* може да бъде обект на измерване по същия начин, по който се измерва *потвърждението*, ако за нея се предложи достатъчно ясна и точна формална експликация. С помощта на такава експликация би трябвало да могат да бъдат сравнявани и съревноваващи се хипотези по тяхната обяснителна сила. Под “съревноваващи се хипотези” в дисертацията се имат предвид несъвместимите хипотези, обясняващи едни и същи свидетелства.

Дисертацията представя основната част от предлаганите досега директни бейсиански мерки на обяснителната сила (Таблица 2).

Таблица 2. Директни бейсиански мерки на обяснителната сила

Мярка на обяснителната сила	Източник
$E_1(e, h) = P(e h) - P(e)$	Eells, 1982; Jeffrey, 1992
$E_2(e, h) = P(e h) - P(e \neg h)$	Nozick, 1981; Christensen, 1999;
$E_3(e, h) = \frac{P(e h) - P(e)}{P(e h) + P(e)}$	Popper, 1935/2005;
$E_4(e, h) = \log \left[\frac{P(e h)}{P(e)} \right]$	Good, 1984;
$E_5(e, h) = \frac{P(h e) - P(h \neg e)}{P(h e) + P(h \neg e)}$	Schupbach and Sprenger, 2011;
$E_6(e, h) = \begin{cases} \frac{P(e h) - P(e)}{1 - P(e)} & \text{ако } P(e h) \geq P(e) \\ \frac{P(e h) - P(e)}{P(e)} & \text{ако } P(e h) < P(e) \end{cases}$	Crupi and Tentori, 2012; 2013;

Обединяваща идея зад директните бейсиански мерки на обяснителната сила е, че обяснителната сила е *нарастваща вероятностна функция* или с други думи, колкото по-

висока стойност на E_i има дадена хипотеза h , толкова *по-добре* тя обяснява наличните свидетелства e (т.е. толкова по-голяма е нейната *обяснителна сила*). Дисертацията ясно показва, че почти всички мерки на обяснителната сила имат за основни компоненти *очакваността на свидетелствата (likelihood)*, т.е. $P(e|h)$, и общата вероятност на свидетелствата, т.е. $P(e)$, изразяваща общата вероятност на свидетелствата изобщо.

3.1. Емпирични изследвания

Настоящата глава обръща внимание на това, че измерването на обяснителната сила има както *нормативна*, така и *дескриптивна* страна. Нормативната страна касае въпроси като: как рационално да мерим обяснителната сила; има ли най-добра мярка на обяснителната сила, приложима във всички случаи; какви основания имаме за използването на дадена мярка на обяснителната сила и др. Измерването на обяснителната сила обаче е и *емпиричен проблем*. В този смисъл, дори когато дадена мярка на обяснителната сила е защитена добре откъм *нормативната* си страна, тя може да бъде критикувана, че не описва адекватно реалното поведение на хората, даващи и получаващи обяснения, т.е. че не е *дескриптивно* адекватна. С оглед на това, някои от водещите философи в теория на потвърдението, в сътрудничество с когнитивни учени, провеждат емпирични изследвания за да установят доколко дескриптивно адекватни са съществуващите бейсиански мерки на потвърдението и обяснителната сила. Един от основните резултати, които обобщават множеството от изследванията на обяснителната сила е, че *съображенията относно качеството на обяснението (explanatory considerations* или ЕС) изглежда играят значима роля в правенето на преценки относно *потвърдението* (Lombrozo, 2012).

Дисертацията също така показва, че емпиричните изследвания, сравняващи дескриптивната адекватност на различните мерки, са все още твърде малко, за да може да бъде направено категорично заключение относно това коя от тези мерки е най-адекватна (и в кои случаи).

3.2. Критики

Тази част от настоящата глава представя критиките на Кларк Глимур (Glymour, 2015) относно бейсианските мерки на обяснителната сила. Първата от тях е, че авторите на въпросните мерки не очертават ясен обхват на приложението им. На второ място, практическата приложимост на много от мерките е под въпрос. Това е така, защото те включват параметрите $P(h|\neg e)$, $P(\neg e|h)$, $P(e)$, $P(\neg e)$, $P(\neg h)$, $P(e|\neg h)$, чиито

стойности са практически трудно- или неустановими. На трето място, Глимур поставя следния проблем: историята на науката е пълна с много добри за времето си обяснения, за които сега знаем, че са погрешни. Ако припишем субективни вероятности (с каквито мерките на обяснителна сила боравят) на тези обяснения от днешна гледна точка, стойностите на някои от мерките, като например E_5 и E_6 (вж. Таблица 2), стават неопределени. На четвърто място в критиката на Глимур е това, че повечето мерки не могат да отразят основни интуиции относно каузалната връзка между експлананс и експланандум. На пето място, дескриптивната адекватност на наличните мерки на обяснителна сила продължава да бъде под въпрос. На шесто и последно място, една добра мярка за обяснителна сила би следвало да може да послужи като вид статистически тест, който да помага в избирането на най-добрата сред наличните обяснителни хипотези. Съществуващите мерки обаче не се предлагат като основа на някакъв вид статистически тест, нито участват в някакъв друг метод за избиране измежду съревноваващи се хипотези. На базата на тази критика, както и на всички останали критики изброени дотук, Глимур прави доста силното обобщение, че досега предложените мерки на обяснителната сила не са сериозни предложения, нито от научна гледна точка, нито от философска такава (Glymour, 2015).

Към горните критики дисертацията добавя още и това, че всички директни мерки на обяснителната сила не взимат предвид приорната вероятност на дадена хипотеза $P(h)$, а са зависими само от нейния *likelihood* $P(e|h)$ и от общата вероятност на свидетелствата $P(e)$. В следващата глава на дисертацията са изложени и други причини да смятаме, че между приорната вероятност и обяснителната сила на една хипотеза съществува връзка, пренебрегването на която води до контраинтуитивни заключения.

4. Обобщение: съвместимостта на бейсианската теория на потвърждението и извода в полза на най-доброто обяснение

Дисертацията прави обобщението, че всички предложения в дебата относно съвместимостта между ВСТ и ИВЕ са съсредоточени върху интеграцията на ВСТ и ИВЕ на ниво подход. Проблемът е там, че съществуват множество различни модели на ВСТ и ИВЕ, като някои от тези модели могат да се окажат съвместими, а други – не. По тази причина в следващата глава на дисертацията е направено ново предложение, което не цели директната *интеграция* на ВСТ и ИВЕ, а търси подходяща бейсианска *експликация* на ключовото за ИВЕ понятие „най-добро обяснение“.

IV. Ролята на приорните вероятности в оценяването на обяснителната сила

Тази глава от дисертацията се фокусира върху два основни проблема, касаещи “извода в полза на най-доброто обяснение” (IBE) и връзката му с бейсианската теория на потвърждението (ВСТ). Първият проблем е липсата на ясна експликация на понятието “най-добро обяснение”. Въпросното понятие е ключово за IBE – ако трябва да се направи извод в полза на *най-доброто* от множество налични съревноваващи се обяснения, то трябва да има и някакъв механизъм, по който да бъдат *сравнени* тези обяснения.

Вторият проблем е съвместимостта между ВСТ и IBE. Както инкомпатибилистите, така и компатибилистите често разглеждат проблема като дихотомия – или ВСТ и IBE са *напълно съвместими*, или те са *напълно несъвместими*. Основен недостатък на такъв вид гледища е, че те не отчитат факта, че има множество *различни* модели на ВСТ и IBE. Някои от тези модели могат да се окажат съвместими, но съвместимостта им може да зависи от *определени условия*. В този смисъл, необходим е метод, чрез който може да бъде изследвано кои са необходимите и/или достатъчни условия, при които два модела, предлагани в рамките на тези различни подходи са съвместими.

За разрешаването на последния проблем, дисертацията предлага да се намери добра вероятностна експликация на понятието “най-добро обяснение”. С други думи, трябва да се намери вероятностен механизъм за *сравняване* на съревноваващи се хипотези по тяхната *обяснителна сила*. По този начин, на въпроса кое е най-доброто от съревноваващите се обяснения е отговорено така: “най-добро е онова обяснение, което има най-висока обяснителна сила по дадената вероятностна мярка на обяснителната сила”. Проблемът тук е, че всички директни бейсиански мерки на обяснителната сила имат общ недостатък в лицето на това, че не отразяват първоначалната степен на убеденост относно разглежданото обяснение (т.е. приорната му вероятност). Има добри причини да се твърди обаче, че приорните вероятности са важни за обяснителната сила и това може да се види ясно в определени случаи. Дисертацията илюстрира това с няколко подобрени примера.

1. Фиктивни примери

Важната роля на приорните вероятности в обяснителната сила може да бъде установена чрез фиктивни примери (*toy examples*), разчитащи на философски интуиции. Един такъв пример е следният:

Представете си, че се събуждате сутринта и виждате, че навън тревата е мокра. (Приемаме твърдението за това наблюдение за твърдение относно наличието на конкретни емпирични свидетелства, т.е. e .) Веднага след това наблюдение формирате две хипотези, обясняващи наблюдавания факт:

- h_1 : “През нощта е валил дъжд.”
- h_2 : “Градинарят е полял тревата рано сутринта.”

Тези две хипотези имат една и съща *очакваност на свидетелствата* (*likelihood*), защото както от h_1 , така и от h_2 следва e . Тук бива въведен обаче и елемент на предварително убеждение относно хипотезите. „Представете си, че в момента е месец юли и че живеете на място, в което дъждовете през месец юли са изключително рядък феномен.“ На базата на тази информация, на h_1 ще бъде приписана доста ниска приорна вероятност, отразяваща убеждението, че валежите през юли на тази географска локация са много редки. В конкретния случай, интуитивно изглежда, че h_2 е много по-добро обяснение на e от h_1 . С други думи, наблюдението, че тревата навън е мокра, е много по-добре обяснено от хипотетичната възможност градинарят да я е полял, отколкото от крайно невероятното обяснение, че през нощта е валил дъжд. Макар че *likelihoods* на хипотезите са равни $P(e|h_1) = P(e|h_2)$, хипотезата с по-висока приорна вероятност $P(h_2) > P(h_1)$ е по-доброто обяснение на свидетелствата. За да отчете подобна интуиция, една бейсианска мярка на обяснителната сила трябва да се влияе от приорните вероятности на сравняваните хипотези (т.е. от $P(h_i)$). Съществуващите директни бейсиански мерки на обяснителната сила, които се влияят само от $P(e|h_i)$ и $P(e)$, не биха могли да отчетат тази интуиция. В конкретния случай те биха достигнали до *контраинтуитивното* заключение, че и двете хипотези *еднакво* добре обясняват направеното наблюдение.

С горния пример дисертацията показва, че разлики в приорните вероятности могат да играят роля в оценяването на обяснителната сила, когато *likelihoods* на сравняваните хипотези са равни. Аргументът обаче бива развит в следващ пример, който показва, че приорните вероятности могат да играят ключова за обяснителната сила роля, дори когато *likelihoods* нямат равни стойности.

Пациент X (45 годишен) има пареза. (Това наблюдение взимаме за нашето свидетелство или e .) Парезата може да е резултат от различни медицински състояния, но с цел опростяването на примера ще вземем предвид само две възможни обяснения:

- h_1 : “X е имал нелекуван сифилис.”
- h_2 : “X е претърпял инсулт.”

Проверявайки медицинското досие на пациента, както и общата статистика на предполагаемите диагнози, лекарката на пациент X има още следната информация:

- X е бил диагностициран със сифилис, но не се е подложил на лечение. (На базата на тази информация, тя приписва много висока приорна вероятност на хипотезата, че X е имал нелекуван сифилис: $P(h_1) = 0.9$.)
- Един от четири пациента, т.е. около 25% от тези, които са с нелекуван сифилис, развиват пареза с напредването на заболяването. (На базата на тази информация, тя може да установи доколко нейното наблюдение на пареза, e , може да бъде очаквано, ако хипотезата, че X има нелекуван сифилис, е вярна: $P(e|h_1) = 0.25$.)
- Четири от пет пациента, т.е. 80%, от претърпелите инсулт получават някакъв вид пареза. (На базата на тази информация, тя може да установи доколко нейното наблюдение на пареза, e , може да бъде очаквано, ако хипотезата, че X е претърпял инсулт, е вярна: $P(e|h_2) = 0.8$.)
- Лекарката не знае дали X е претърпял инсулт, но знае, че в неговата възрастова група между двама и четирима души на хиляда, т.е. 0.2% ~ 0.4%, са с висок риск за инсулт. (На базата на тази информация тя приписва доста ниска приорна вероятност на хипотезата, че X е претърпял инсулт: $P(h_2) = 0.004$.)

В този случай, макар че $P(e|h_2) > P(e|h_1)$, повечето лекари, при наличието на информация i) – iv), биха приписали по-висока обяснителна сила на h_1 . С други думи, макар че парезата е много по-очаквана след инсулт, отколкото при нелекуван сифилис, повечето лекари биха обяснили наличието на пареза с медицинската история на нелекуван сифилис. Интуитивно, h_1 е по-доброто обяснение на e от h_2 , макар че h_2 има по-висок *likelihood*.

Директните бейсиански мерки на обяснителната сила обаче биха били принудени да направят обратното заключение. Те биха взели предвид само $P(e|h_2) > P(e|h_1)$ и биха заключили, h_2 е по-доброто обяснение на свидетелствата от h_1 , което в случая е контраинтуитивно.

2. Примери от научната практика

Основната функция на двата примера от науката, разгледани в дисертацията, е да покажат, че ако в бейсианската експликация на понятието “най-добро обяснение” вземем предвид *и* *приорните вероятности* на оценяваните хипотези, а не само техните *likelihoods*, то можем да достигнем до същия *извод в полза на най-доброто обяснение*, до който достигат и учените от съответния пример. С помощта на тези примери е демонстрирано и как *не бихме достигнали до същите изводи* като съответните експерти в дадена област, ако използваме *единствено likelihoods* в определянето на това коя е най-добрата обяснителна хипотеза.

2.1. Алфред Вегенер и теорията за континенталния дрейф

Първият пример от историята на науката е с т.нар. “теория за континенталния дрейф” на немския метеоролог и климатолог Алфред Вегенер. През 1912 година, Вегенер предлага хипотетичното твърдение, че Земяните континенти не са статични, а са в постоянно движение (наподобяващо дрейф), осъществявано по дъното на световния океан. Ако проследим това движение назад в историята до около 200 милиона години, то ще видим, че днешните континенти са образували в началото един единствен праисторически суперконтинент (наречен “Пангея”), който впоследствие се е разпаднал, под въздействието на силите, причинили и съпътстващи движението на този континент (Frisch, Meschede and Blakey, 2011; Frankel, 2012).

С времето Вегенер развива тази своя хипотеза, като с нея обяснява няколко наблюдения:

- Контурите на съвременните континенти (особено на нивото на техния континентален шелф) изглежда пасват един на друг подозрително добре.
- Когато съединим континентите по пасващите им контури, други техни характеристики също си пасват добре. Геоложките пластове от Апалачите в източната част на Северна Америка съвпадат с тези в Шотландия. Вкаменелости от едни и същи животински и растителни видове могат да бъдат намерени по крайбрежията на континенти, които си пасват по контури, но днес са разделени от огромни океани.
- Вкаменелости от определени видове се намират на континенти, на които никога не би могло да има подобни видове, предвид историята на климата им. В скалите на Аляска например са намерени вкаменелости на субтропически растения (Chaney, 1940).

Доминиращата теория по това време, на която хипотезата за континенталния дрейф се явява алтернатива, твърди, че континентите са стабилни и статични – те не се движат и са били на позициите, на които са и днес, в продължение на милиони години. За да обясни наличието на вкаменелости от едни и същи видове на различните континенти, тази алтернативна теория предполага наличието на огромни мостове от земна маса между континентите, които с течение на времето са потънали. Макар да може да обясни наличието на вкаменелости от *мигриращи* видове на различните континенти с това, че въпросните видове са прекосили съществуващите тогава земни мостове, тази теория не може да обясни наличието на вкаменелости от едни и същи *растителни* видове на различните континенти. Тя не може да обясни и други свидетелства, като например съвпадения в характеристиките на ледниците или съвпадения в геоложките пластове между континентите.

Когато обаче Вегенер публикува хипотезата си през 1912, тя е посрещната меко казано негативно: тя е подложена на унищожителна критика, а самият Вегенер е взет на присмех. Няколко фактора допринасят за този прием на хипотезата за континенталния дрейф. Първият и най-важен от тях е, че Вегенер не успява да посочи правдоподобен физически механизъм за движението на континентите. Вторият фактор за “хладния” прием на хипотезата за континенталния дрейф е, че изчисленията на Вегенер дават неточна стойност на скоростта на континенталния дрейф. Третият фактор несъмнено е гледището, споделяно от мнозина геолози, че *метеорологът* Вегенер не е специалист в областта на геологията.

Съществува обаче и малцинство от геолози, които подкрепят хипотезата и продължават да търсят свидетелства за нея. Сред тях са: геологът Александър дю Тоа от Южна Африка, който намира свидетелства за континенталния дрейф в приликата между геоложките пластове на Южна Африка и Южна Америка; швейцарският геолог Емил Арганд, който вижда свидетелства за континенталния дрейф в образуването на Швейцарските Алпи; австралийският геолог Самюел Кери, който по-късно става поддръжник на хипотезата, че движението на континентите е причинено от разширяването на обема на Земята; и британският геолог Артър Холмс, който предлага физически механизъм за континенталния дрейф, на мястото на неубедителния механизъм на Вегенер (Frankel, 2012).

Хипотезата на Вегенер за континенталния дрейф на практика се явява пряк предшественик на съвременната теория за тектониката на плочите. Въпреки разликите

между съвременната теория и Вегенеровата хипотеза, от днешна гледна точка Вегенер е бил, ако не изцяло прав, то със сигурност на път в правилната посока. В този смисъл, неговото предложение не просто не е било за осмиване, а е заслужавало доста сериозно внимание. В отговор на това защо то не е получило такова внимание, дисертацията предлага анализ в рамките на теория на потвърждението.

Когато Вегенер публикува своята хипотеза за континенталния дрейф през 1912 година (и по-късно в книга през 1915 година), тя се превръща в алтернатива на вече съществуващата теория за статичните континенти. В дисертацията са експлицирани двете съревноваващи се хипотези:

- h_1 – хипотезата на Вегенер за континенталния дрейф;
- h_2 – хипотезата за статичните континенти.

Двете хипотези имат за цел да обяснят едно и също множество от емпирични свидетелства (e). В e влизат: (e_1) намерените вкаменелости от едни и същи растителни и животински видове по различните континенти, (e_2) наличието на вкаменелости от растителни видове на континенти с напълно непригоден за тях климат, (e_3) съвпадащите характеристики на някои геоложки пластове, както и на някои ледници, намиращи се на различни континенти, (e_4) съвпадащите очертания на крайбрежните ивици на континентите; тук влизат още и ($\neg e_5$) липсата на следи по океанското дъно от преминаването през него на континенти, а също и ($\neg e_6$) липсата на следи потъването на огромни земни маси, свързващи континентите.

Приемайки широко разпространена във философия на науката интуиция, че дадена хипотеза обяснява наличните свидетелства, ако ги прави по-очаквани, дисертацията показва, че двете хипотези обясняват свидетелствата e до различна степен. Вегенеровата хипотеза за континенталния дрейф (h_1) прави по-очаквани всички свидетелства от множеството, с изключение на $\neg e_5$, на базата на основното ѝ допускане, че континентите са изменяли позициите си по земната повърхност с течение на времето. Хипотезата за статичните континенти (h_2) прави по-очаквани e_1 и e_2 , но само чрез допълнителни допускания. Също така, h_2 не може да направи e_3 или e_4 по-очаквани, обявявайки ги просто за съвпадения, но може да направи $\neg e_5$ напълно очаквано на базата на основната си предпоставка – ако континентите наистина винаги са били статични, то е напълно в реда на нещата да не открием следи от движението им по океанското дъно. Тя обаче не

може да направи по-очаквано и $\neg e_6$, макар че защитниците ѝ могат да пледират за пощателно издирване на свидетелства за потъналите земни мостове.

Дисертацията показва, че h_1 прави по-очаквани повече свидетелства от множеството e от h_2 , като също така прави по-очаквани e_1 и e_2 без да са ѝ необходими *ad hoc* допускания. Очакваността на свидетелствата на двете хипотези (*likelihoods*) биват представени по следния начин: $P(e|h_1) > P(e|h_2)$. Казано иначе, ако хипотезата на Вегенер за континенталния дрейф е истинна, това би направило наличните свидетелства по-очаквани, отколкото ако хипотезата за статичните континенти е истинна. В дисертацията се твърди още, че ако за експликация на понятието “най-добро обяснение” бъде приета мярка на обяснителната сила, която взема предвид само *likelihoods*, то неизбежен става изводът, че хипотезата за континенталния дрейф е най-доброто обяснение на свидетелствата, поне в сравнение с хипотезата за статичните континенти. Макар този извод да е правилен от историческа гледна точка, той не обяснява защо учените от онова време отхвърлят Вегенеровата хипотеза така категорично. Очевидно в техните разсъждения относно това кое е най-доброто обяснение влизат някакви допълнителни фактори, отвъд просто очакваността на свидетелствата. В дисертацията е предложено тези фактори да бъдат отчетени с помощта на приорните вероятности.

В изложението на примера вече бе показано, че основният недостатък на хипотезата за континенталния дрейф, за който е критикуван Вегенер, е неубедителният физически механизъм, с който той опитва да обясни движението на континентите. Съвсем естествено, това намалява драстично приорната вероятност на Вегенеровата хипотеза. Хипотеза, която разчита на физически неправдоподобен механизъм и не предлага алтернатива, не е особено убедителна на първо място. Още повече, настоящата глава показва, че към хипотезата за континенталния дрейф има съмнения и поради това, че самият Вегенер не е специалист в областта на геологията. Вегенер е метеоролог с докторат по астрономия, което в очите на мнозина геолози от онова време веднага прави съмнителна всякаква геофизическа хипотеза, която Вегенер би опитал да им предложи. Макар да не е най-рационалният довод за съмнение в научната издържаност на предлаганата от Вегенер хипотеза, то със сигурност подобен предразсъдък би бил фактор при формирането на приорната ѝ вероятност. В допълнение, Вегенер дава неточни стойности в изчисленията си на скоростите на раздалечаване на континентите. Тези стойности са в някои случаи толкова завишени, че предизвикват скептицизма на геофизиците дори преди наличието на конкретни емпирични свидетелства.

Всички тези фактори драстично намаляват приорната вероятност на хипотезата на Вегенер в очите на геолог от онова време и то до такава степен, че ниският ѝ приор неутрализира преимуществото, която тя има откъм по-висок *likelihood*. В този смисъл, описанието на историческия случай ясно показва защо за повечето геолози по времето на Вегенер хипотезата за континенталния дрейф не е най-доброто обяснение на наличните свидетелства и те предпочитат вече установената хипотеза за статичността на континентите. За достигането до този извод обаче, трябва да бъдат взети предвид въпросните приорни вероятности на двете съревноваващи се хипотези. Ако вместо това решението относно най-доброто обяснение е базирано единствено върху *likelihoods*, въпросът защо хипотезата за континенталния дрейф първоначално получава толкова отрицателен прием от обществото на геолозите, би останал без отговор.

2.2. Батигин и Браун и хипотезата за „Планета девет“

Вторият пример в настоящата глава касае хипотетичното съществуване на девета, досега ненаблюдавана планета в нашата Слънчева система, направено от двама астрономи от Калифорнийския технологичен институт (*Caltech*) – Константин Батигин и Майкъл Браун (Batygin and Brown, 2016). Тази хипотеза бива предложена като най-доброто обяснение на някои странни характеристики на данните от наблюдението на шест обекта отвъд орбитата на Нептун. Въпросните шест обекта са от разредния диск и пояса на Кайпер (“обекти от пояса на Кайпер”, оттук нататък ОПК), като перихелиите на техните орбити са по-големи от орбитата на Нептун, а големите им полуоси са по-големи от 150 астрономически единици ($a > 150 \text{ AU}$). Странната характеристика в орбитите на тези обекти е скупчването на параметрите на перихелия им около 0° ($\omega \approx 0^\circ$). С други думи, перихелият на всеки от тези шест обекта лежи на еклиптичната равнина, а възходящите им възли съвпадат с перихелиите, което определя и общата посока на техните орбити – от юг на север.

Батигин и Браун изчисляват, че за орбити с $a > 50 \text{ AU}$, подобна характеристика се проявява само в 0.007% от случаите. Вероятността е толкова ниска, защото с течение на времето скупчените параметри на перихелия имат тенденцията да се разместват. Нашата Слънчева система съществува от достатъчно много време, така че параметрите на перихелия на шестте ОПК трябва вече да са се разместили, освен ако не са придържани от някакъв физически механизъм. Съществуването на хипотетичната ненаблюдавана

планета Батигин и Браун предлагат именно като основа на такъв механизъм, който държи перихелиите на шестте ОПК осцилиращи около $\omega = 0^\circ$.

За аномалията в орбитите на шестте ОПК в дисертацията са предложени още две съревноваващи се обяснения. Първото е на Трухило и Шепард (Trujillo and Sheppard, 2014), които показват, че обект с маса около пет пъти по-голяма от Земята би могъл да придържа периметъра на перихелия на даден ОПК около $\omega = 0^\circ$ чрез физически механизъм, известен като “резонанс на Лидов – Козай” (или просто “механизъм на Козай”). Въпросният механизъм описва влиянието на гравитацията на обект с външна орбита върху друг обект, чиято орбита се намира вътре в орбитата на първия.

Второто обяснение е на Мадиган и МакКорт (Madigan and McCourt, 2015), според които скупчването на параметрите на перихелия на ОПК обектите е породено от гравитационна нестабилност в системата на самия разреден диск на пояса на Кайпер. Гравитационната нестабилност, породена от взаимодействието на различните части на една и съща се нарича “самогравитационна нестабилност” (*self-gravitational instability*). Самогравитационната нестабилност на Мадиган и МакКорт е вид нестабилност на инклинацията, т.е. нестабилност в наклона на орбитите на телата в дадена гравитационна система, която постепенно изменя формата на системата от диск в конус, като събира параметрите на перихелия на орбитиращите тела и експоненциално увеличава тяхната инклинация (Madigan and McCourt, 2015). В изложението на примера, акцентът е поставен върху това, че ако разредният диск и поясът на Кайпер притежават достатъчна маса, за да задействат гравитационните процеси на въпросния механизъм на нестабилност, това би обяснило скупчването на параметрите на перихелия на шестте ОПК.

Настоящата глава излага и основната критика към съревноваващите се обяснителни хипотези. Относно хипотезата на Трухило и Шепард: за да може механизмът на Козай да обясни скупчването на перихелиите на всичките шест ОПК около $\omega = 0^\circ$, той се нуждае от предполагаемото съществуване на *няколко* ненаблюдавани досега масивни тела с *точно определени орбити*. Още повече, че сам по себе си механизмът на Козай не може да обясни защо са наблюдавани обекти с перихелии, скупчени около $\omega = 0^\circ$, но не е наблюдавано подобно скупчване около $\omega = 180^\circ$ (Batygin and Brown, 2016). За да може да обясни липсата на подобно наблюдение, предлаганият от Трухило и Шепард механизъм изисква силното влияние на външна звезда в миналото на Слънчевата система, но подобно предположение не съвпада с това, което знаем за Слънчевата

система досега. Тези две предпоставки – наличието на множество ненаблюдавани масивни тела и на силно външно гравитационно влияние върху Слънчевата система, породено от близка среща с друга звезда, като цяло правят обяснението на Трухило и Шепард по- *ad hoc*.

Що се отнася до второто обяснение, това на Мадиган и МакКорт, предложението от тях механизъм на самогравитационна нестабилност се нуждае от предпоставката, че разредният диск и пояса на Кайпер са били много по-масивни в миналото, отколкото съвременните изчисления предполагат. Само така те биха могли да имат достатъчно гравитация, за да сработи въпросният механизъм. Проблемите с това предложение са два. От една страна, за момента няма свидетелства за подобна предпоставка. От друга, предпоставката е проблемна и от теоретична гледна точка – по-голямата част от масата, която разредният диск и поясът на Кайпер може и да са имали в миналото, е твърде вероятно да е била изхвърлена извън Слънчевата система при взаимодействия с масивните планети на външната Слънчева система – газовите гиганти Юпитер и Сатурн или ледените гиганти Уран и Нептун..

На фона на недостатъците на горните две алтернативни обяснения, за най-добро обяснение на наблюдаваното скупчване на перихелиите на шестте ОПК е определено наличието на обект с маса около 10 Земни маси, в орбита около Слънцето с голяма полуос $a \approx 700$ AU, перихелий около 200 AU и афелий (най-далечна от Слънцето точка на орбитата) около 1200 AU (Batygin and Brown, 2016). Предполага се, че този обект, с работно наименование „Планета девет“ (*Planet Nine*), най-вероятно е леден гигант, формиран от голямо планетарно ядро, изхвърлено от вътрешността на Слънчевата система през началните фази на формирането ѝ. Орбиталният му период вероятно е между 10 и 20 хиляди години, като през повечето от това време той е твърде далече от Земята, за да може да бъде наблюдаван без помощта на инструменти с изключително висока резолюция. Тази му отдалеченост, особено съчетана с ниско алbedo (т.е. ниска отражателна способност на обекта), може да обясни защо той все още не е наблюдаван.

В обобщение, настоящата глава представя три съревноваващи се обяснителни хипотези, от всяка от които следват емпиричните свидетелства, т.е. наблюдаваното скупчване на перихелиите на шестте ОПК. Това са:

- h_1 – хипотезата на Трухило и Шепард (Trujillo and Shepherd, 2014), чийто модел изисква като предпоставка съществуването на няколко неоткрити досега масивни обекта;
- h_2 – хипотезата на Мадиган и МакКорт (Madigan and McCourt, 2015), която изисква като предпоставка разредния диск на пояса на Кайпер да е бил по-масивен от това, което сочат днешните изчисления;
- h_3 – хипотезата на Батигин и Браун (Batygin and Brown, 2016), която предполага съществуването на “Планета девет”.

На базата на тези обяснителните хипотези, дисертацията показва, че ако за интерпретация на понятието “най-добро обяснение” в ИВЕ бъде приета някоя от мерките на обяснителната сила, които взимат предвид само *likelihood*, то неизбежно следва заключението, че горните три хипотези са еднакво добри обяснения на свидетелствата. Това е така, защото и от трите хипотези свидетелствата следват директно, което означава, че *likelihoods* на трите хипотези са равни: $P(e|h_1) = P(e|h_2) = P(e|h_3) \approx 1$, където e е твърдение за наличието на наблюдаваните свидетелства (скупването на перихелиите на шестте ОПК). Директните бейсиански мерки на обяснителната сила се влияят само от $P(e|h_i)$ и $P(e)$, като и за трите хипотези стойностите на тези параметри в случая са равни. Изводът, че трите хипотези имат еднаква обяснителна сила, обаче би противоречал на експертното мнение на онези астрономи, които смятат, че хипотезата, предполагаща съществуването на ненаблюдавана планета, е по-доброто обяснение на свидетелствата. В тази насока са приведени свидетелства от изказванията на астрономите Родни Гомес и Алесандро Морбидели.

Дисертацията показва, че проблемът би бил разрешен, ако във формалната експликация на понятието “най-добро обяснение” бъдат отчетени разликите в приорните вероятности между съревноваващите се хипотези. Хипотезата на Трухило и Шепард (h_1) изисква наличието на *няколко* масивни ненаблюдавани досега тела, докато тази на Батигин и Браун (h_3) изисква наличието само на *един* такъв масивен обект. Според теория на вероятностите, вероятността да съществува един ненаблюдаван масивен обект винаги ще е по-висока от вероятността да съществуват няколко такива обекта едновременно. В този смисъл, хипотезата, която предполага съществуването на един ненаблюдаван масивен обект в Слънчевата система (h_3) трябва да получи по-висока приорна вероятност от хипотезата, която предполага съществуването на няколко такива обекти (h_1), т.е. $P(h_3) > P(h_1)$. В противен случай, приписаните стойности на приорите не биха

били нито рационални, нито консистентни (в смисъла на спазващи аксиомите на теория на вероятностите).

Хипотезата на Мадиган и МакКорт (h_2) пък изисква предпоставката, че разредният диск на пояса на Кайпер е бил много по-масивен от това, което сочат днешните изчисления. В допълнение, тази предпоставка е малко вероятна и поради теоретични причини – ако тази маса изобщо е съществувала, тя трябва относително бързо да е била изхвърлена от системата след взаимодействия с газовите и ледените гиганти. При равни други условия, една хипотеза (като h_2), която не е в съответствие с наличните изчисления и освен това е слабо вероятна от теоретична гледна точка, не би получила по-висока приорна вероятност от друга хипотеза, която не се сблъсква с подобни проблеми (като h_3). В този смисъл, дисертацията прави извода, че $P(h_3) > P(h_2)$.

Оттук бива изведено заключението, че хипотезата на Батигин и Браун има най-високата приорна вероятност от трите съревноваващи се хипотези: $P(h_3) > P(h_1)$ и $P(h_3) > P(h_2)$. Взимайки предвид този извод, в решенията относно това коя от трите хипотези е най-доброто обяснение на свидетелствата, би се достигнало до заключение, съответстващо на мнението на експертите в областта. С други думи, приемането за адекватна експликация на понятието “най-добро обяснение” някаква бейсианска мярка на обяснителната сила, която взема предвид и приорните вероятности, избягва контраинтуитивното заключение, че съревноваващите се обяснителни хипотези имат еднаква обяснителна сила.

3. Метод за изследване на условията на съвместимост между бейсианската теория на потвърдението и извода в полза на най-доброто обяснение

Дисертацията представя още едно предимство на преследването на адекватна бейсианска експликация на понятието “най-добро обяснение”: това, че тя позволява изследването на условията на съвместимост между ВСТ и ИВЕ по директен начин. Под “съвместимост” между двата подхода към потвърдението в случая се разбира възможността им да дават едни и същи резултати, когато бъдат приложени към един и същи случай. Както вече бе отбелязано в началото на тази глава, съвместимостта между двата подхода зависи от конкретните модели на ВСТ и ИВЕ, които биват сравнявани. В допълнение, съвместимостта между два модела вероятно може да бъде установена само при

определени условия. За установяването на тези условия е предложен формален метод, който е изложен накратко и тук.

Ако за вероятностна експликация на “най-добро обяснение” бъде приета следната мярка на обяснителната сила:

$$E_7(e, h) = P(e|h) \times P(h) \quad (1)$$

С други думи, за най-добро обяснение, от множество съревноваващи се обяснителни хипотези, бива прието това, което има най-висока стойност по E_7 . Оттук, ИВЕ бива интерпретирано като правилото, според което на *най-добрата обяснителна хипотеза* следва да се припише и *най-висока степен на потвърждение*.

Мярката E_7 бива приета само заради примера, и не трябва да бъде приемана като сериозно предложение за мярка на обяснителната сила.

Чрез E_7 бива дефинирана обяснителната сила на две съревноваващи се хипотези h_1 и h_2 , които обясняват едни и същи емпирични свидетелства e :

$$E_7(e, h_1) = P(e|h_1) \times P(h_1) \quad (2)$$

$$E_7(e, h_2) = P(e|h_2) \times P(h_2) \quad (3)$$

Това, което се търси в случая са условията, при които гореизложената интерпретация на ИВЕ може да бъде съвместима с мярката на потвърждение предложена от Eells (1982) и защитавана от Jeffrey (1992):

$$C_1(e, h) = P(h|e) - P(h) \quad (4)$$

Според тази мярка, потвърждението е нарастваща функция на разликата между постериорната и приорната вероятност на хипотезата. Чрез C_1 бива дефинирано потвърждението на разглежданите две хипотези по следния начин:

$$C_1(e, h_1) = P(h_1|e) - P(h_1) \quad (5)$$

$$C_1(e, h_2) = P(h_2|e) - P(h_2) \quad (6)$$

Оттук съвместимостта на ВСТ и ИВЕ в изследвана в сценарий, в който h_1 е по-добре потвърдена от свидетелствата e , отколкото h_2 . Според гореизложената интерпретация на ИВЕ, h_1 е по-добре потвърдена от h_2 , ако е задоволено следното условие:

$$E_7(e, h_1) > E_7(e, h_2) \quad (7)$$

$$P(e|h_1) \times P(h_1) > P(e|h_2) \times P(h_2) \quad (8)$$

Според избраната бейсианска мярка на потвърдението пък, h_1 е по-добре потвърдена от h_2 , ако:

$$C_1(e, h_1) > C_1(e, h_2) \quad (9)$$

$$P(h_1|e) - P(h_1) > P(h_2|e) - P(h_2) \quad (10)$$

Тъй като h_1 и h_2 обясняват едни и същи свидетелства, (8) може да бъде трансформирано като разделим и двете му страни на $P(e)$, ако $P(e) > 0$:

$$\frac{P(e|h_1) \times P(h_1)}{P(e)} > \frac{P(e|h_2) \times P(h_2)}{P(e)} \quad (11)$$

$$P(h_1|e) > P(h_2|e) \quad (12)$$

На свой ред, (10) може да бъде трансформирано по следния начин:

$$P(h_1|e) - P(h_2|e) > P(h_1) - P(h_2) \quad (13)$$

След което е трансформирано и (12) в:

$$P(h_1|e) - P(h_2|e) > 0 \quad (14)$$

При предположението, че:

$$P(h_1) - P(h_2) \geq 0 \quad (15)$$

От (13) и (15) можем да бъде изведено следното заключение:

$$P(h_1|e) - P(h_2|e) > 0 \quad (16)$$

Тъй като (14) и (16) са еквивалентни, (13) и (14) също са еквивалентни, при изпълнено условие (15). Казано иначе, както ИВЕ, така и ВСТ биха заключили, че h_1 е по-добре потвърдено от e от h_2 , ако $P(h_1) \geq P(h_2)$, а когато резултатите им се разминават, това е поради нарушение на условие (15).

Откриването на подобни условия на съвместимост може да има два типа продължение. От една страна, би могло да се търси рационална обосновка на условието. С други думи, могат да бъдат търсени аргументи за това защо даденото условие не е просто странна математическа особеност, а има конкретен смисъл. От друга страна, ако условието се окаже незащитимо или пък силно контраинтуитивно, друг рационален ход на действие е да се потърси различна мярка на обяснителната сила или пък различна мярка на

потвърждението и методът да бъде приложен отново, за да се разбере дали те биха били съвместими, и ако са – при какви условия.

В настоящата глава е отбелязано още, че представеният по-горе метод може да бъде безпроблемно приложен само за интерпретации на IBE, които използват вероятностна експликация на понятието “най-добро обяснение”. В този смисъл, подобни интерпретации на IBE имат предимство пред класическото IBE, защото позволяват изследването на съвместимостта на ВСТ и IBE модел по модел, вместо да се опитват да защитават крайните (но неточни) твърдения, че те са напълно съвместими или пък напълно несъвместими.

4. Обобщение: приорните вероятности и извода в полза на най-доброто обяснение

Настоящата глава на дисертацията показва защо една бейсианска експликация на ключовото за IBE понятие “най-добро обяснение”, която е под формата на вероятностна мярка на обяснителната сила, трябва да взема предвид приорните вероятности на оценяваните хипотези. В сравнение със съществуващите директни мерки на обяснителната сила, такава една мярка би била по-адекватна, когато бъде приложена към фиктивни примери, като тези представени първата част на главата. За разлика от съществуващите мерки за обяснителна сила, такава мярка не би довела до силно контраинтуитивни изводи относно това кое е най-доброто обяснение. Също така, подобна мярка би била дескриптивно по-точна, когато бъде приложена към примери от науката, като представените във втората част на главата. Тя би позволила правенето на извод, сходен с този на учените, относно това коя е най-добрата от съревноваващите се обяснителни хипотези. Директните мерки на обяснителната сила в литературата, които се влияят само от *likelihoods*, не позволяват правенето на такъв извод в полза на най-доброто обяснение.

Бейсианската експликация на понятието “най-добро обяснение”, под формата на адекватна вероятностна мярка на обяснителната сила, позволява също така изследването на съвместимостта на ВСТ и IBE модел по модел, като за всяка двойка модели бъде проверено при какви условия те са съвместими (ако изобщо). За целта на такова изследване в дисертацията е предложен метод за установяване на съвместимост между резултатите, които конкретни модели на ВСТ и IBE дават, когато сравняват едни и същи хипотези.

V. Заключение

Един от основните проблеми, които дисертацията разглежда, е този за съвместимостта между два, на пръв поглед конкурентни подхода към потвърждението: бейсианската теория на потвърждението (*Bayesian confirmation theory* или просто ВСТ) и теорията, основаваща се върху т.нар. „извод в полза на най-доброто обяснение“ (*inference to the best explanation* или ИВЕ). Основното заключение в дисертацията относно този дебат е, че както инкомпатибилизмът, така и компатибилизмът по отношение на двата подхода, за момента се провалят в решаването на основния проблем, който разглеждат. С оглед на това заключение, е предложено различно решение на проблема, което предполага отместването на фокуса от ниво *подход* към ниво *отделен модел*, с цел систематичното изследване на съвместимостта на ВСТ и ИВЕ *модел по модел*. За целта е въведен нов формален метод, който сравнява резултатите от даден модел на ВСТ и даден модел на ИВЕ, когато те бъдат приложени към един и същи случай. С помощта на нововъведения метод могат да бъдат определяни *условията*, които трябва да бъдат изпълнени, за да бъдат получените резултати от двата сравнявани модела еднакви, на базата на което може да бъде обявена и тяхната *условна съвместимост*. За систематичното приложение на новия метод е намерено решение на два други проблема пред ИВЕ. Това са: липсата на задоволителна формална *експликация* на ИВЕ, както и липсата в ИВЕ на формален *механизъм*, чрез който да става изборът на най-добро обяснение.

В тази връзка са разгледани съществуващите в литературата директни бейсиански мерки на обяснителната сила, като възможности за експликация на ключовото за ИВЕ понятие „най-добро обяснение“. При това разглеждане е установено, че тези мерки се влияят само от два параметъра: от това доколко „очаквани“ са емпиричните свидетелства, ако разглежданата хипотеза е истинна (т.нар. *likelihood* или $P(e|h)$) и от общата вероятност на емпиричните свидетелства по принцип (т.е. от $P(e)$). Изолирането на третия основен бейсиански параметър – т.нар. „приорна вероятност“, $P(h)$, отразяваща първоначалното убеждение в истинността на дадено твърдение, хипотеза или теория, поражда съмнение в адекватността на съществуващите мерки. Това поражда нов проблем: доколко може да се разчита на съществуващите директни бейсиански мерки на обяснителната сила за адекватната експликация на ИВЕ-понятието „най-добро обяснение“? В отговор на този въпрос, са разгледани четири примера – два фиктивни и два от науката. Чрез тях е показано, че ако ИВЕ-понятието „най-добро обяснение“ бъде експлицирано с някоя от съществуващите директни бейсиански мерки на обяснителната сила, които се влияят

основно от очакваността на свидетелствата и игнорират приорните вероятности на сравняваните хипотези, то това би довело до два вида проблеми. От една страна, то би довело до контраинтуитивни заключения относно това кое е най-доброто обяснение в хипотетични сценарии като тези, представени в двата фиктивни примера. От друга страна, то би довело до разминавания с мненията на учените относно най-доброто обяснение в някои реални случаи, като например този със студения прием на хипотезата на Алфред Вегнер за континенталния дрейф в първите десетилетия на XX век или пък този с относителната убеденост на някои съвременни астрономи в съществуването на ненаблюдавана „Планета девет“ в покрайнините на Слънчевата система. На базата на тези резултати е направено заключението, че съществуващите директни бейсиански мерки не са адекватни мерки на обяснителната сила.

С помощта на представените примери е направено и още едно наблюдение. Ако в отсъжданията относно най-доброто обяснение бъдат включени и приорни вероятности, формирани на базата на предварителната информация, която е налична във всеки един от примерните случаи, това би направило възможно както достигането до интуитивни заключения във фиктивните примери, така и обяснението на това защо учените поддържат съответното мнение в примерите от науката. В дисертацията, това наблюдение води до заключението, че за експликация на ИВЕ-понятието „най-добро обяснение“ трябва да бъде избрана такава вероятностна мярка на обяснителната сила, която да отчита приорните вероятности. В противен случай експликацията би била дескриптивно неадекватна.

Справка за приносите

Дисертацията отчита следните приноси:

1. Относно въпроса за съвместимостта на двата подхода към потвърждаването на научни хипотези – бейсианската теория на потвърждението (ВСТ) и теорията, основана на извод в полза на най-доброто обяснение (IBE): дисертацията показва, че освен тезите на инкомпатибилизма и компатибилизма е възможно да се защитава и трета позиция, според която двата подхода са съвместими при определени условия.
2. В дисертацията е предложен формален метод за определянето на условията на ВСТ-IBE съвместимост.
3. С цел успешното приложение на формалния метод от т. 2, в дисертацията е предложено ключовото за IBE понятие „най-добро обяснение“ да бъде формално експлицирано под формата на бейсианска мярка за обяснителната сила.
4. Дисертацията показва, че съществуващите директни бейсиански мерки на обяснителната сила не са адекватно средство за експликация на понятието „най-добро обяснение“, тъй като приложението им в различни видове случаи води до незадоволителни резултати. В изкуствени примери, то води до силно контраинтуитивни резултати, а в случаи от реалната научна практика – до резултати, които се разминават с преценките на експертите.
5. В отговор на открития в т. 4 проблем, дисертацията предлага вероятностната експликация на понятието „най-добро обяснение“ да отчита и приорните вероятности на оценяваните хипотези. Това предложение е подкрепено със същите примери, които демонстрират недостатъците на съществуващите директни бейсиански мерки на обяснителната сила.

Публикации по темата на дисертацията

1. Donchev, A., 2017. The Role of Priors in a Probabilistic Account of “Best Explanation”. *Organon F*, 24(4), pp.511-525.
2. Дончев, А., 2016. Положителната роля на субективния фактор в науката: защита срещу Попървата критика на субективистката вероятност. *Философски алтернативи*, 25(2), сс.16-22.
3. Дончев, А., 2015. Основни проблеми пред хипотетико-дедуктивизма и техните вероятностни решения. *Философски алтернативи*, 23(6), сс.33-41.

Използвана литература

1. Barnes, E., 1995. Inference to the Loveliest Explanation. *Synthese*, 103(2), pp.251-277.
2. Batygin, K. and Brown, M., 2016. Evidence for a Distant Giant Planet in the Solar System. *The Astronomical Journal*, 151(22), pp.1-12.
3. Boyd, R., 1980. Scientific Realism and Naturalistic Epistemology. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume Two: Symposia and Invited Papers, pp.613-622.
4. Boyd, R., 1983. On the Current Status of the Issue of Scientific Realism. *Erkenntnis*, 19(1/3), pp.45-90.
5. Boyd, R., 1985. Lex Orandi est Lex Credendi. In: P.M. Churchland and C.A. Hooker, eds. *Images of Science: Essays on Realism and Empiricism*. Chicago, IL: University of Chicago Press. pp.3-34.
6. Carnap, R., 1950/1962. *Logical Foundations of Probability* (2nd ed.). Chicago: University of Chicago Press.
7. Chaney, R., 1940. Bearing of Forests on the Theory of Continental Drift. *The Scientific Monthly*, 51(6), pp.489-499.
8. Crupi, V., 2013. Confirmation. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta, ed. [online] Available at: [<https://plato.stanford.edu/archives/win2016/entries/confirmation/>](https://plato.stanford.edu/archives/win2016/entries/confirmation/) [Accessed 30 January 2018].
9. Crupi, V. and Tentori, K., 2012. A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems). *Philosophy of Science*, 79, pp.365-385.
10. Crupi, V. and Tentori, K., 2013. Confirmation as Partial Entailment: A Representation Theorem in Inductive Logic. *Journal of Applied Logic*, 11, pp.364-372.
11. Crupi, V., Tentori, K. and Gonzalez, M., 2007. On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues. *Philosophy of Science*, 74(2), pp.229-252.
12. Douven, I., 2011. Abduction. *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta, ed. [online] Available at: [<http://plato.stanford.edu/archives/spr2011/entries/abduction/>](http://plato.stanford.edu/archives/spr2011/entries/abduction/) [Accessed 01 February 2018].

13. Duhem, P., 1906/1954. *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.
14. Eells, E., 1982. *Rational Decision and Causality*. Cambridge: Cambridge University Press.
15. Frankel, H., 2012. *The Continental Drift Controversy. Volume IV: Evolution into Plate Tectonics*. Cambridge: Cambridge University Press.
16. Frisch, W., Meschede, M. and Blakey, R., 2011. *Plate Tectonics: Continental Drift and Mountain Building*. Berlin: Springer.
17. Gillies, D., 2000. *Philosophical Theories of Probability*. London: Routledge.
18. Glymour, C., 1980. Hypothetico-Deductivism Is Hopeless. *Philosophy of Science*, 47(2), pp.322-325.
19. Glymour, C., 2015. Probability and the Explanatory Virtues. *British Journal for the Philosophy of Science*, 66(3), pp.591-604.
20. Good, I.J., 1984. The Best Explicatum for Weight of Evidence. *Journal of Statistical Computation and Simulation*, 19, pp.294-299.
21. Goodman, N., 1983. *Fact, Fiction and Forecast* (4th ed.). Cambridge (MA): Harvard University Press.
22. Hempel, C., 1945a. Studies in the Logic of Confirmation (I). *Mind*, New Series, 54(213), pp.1-26.
23. Hempel, C., 1945b. Studies in the Logic of Confirmation (II). *Mind*, New Series, 54(214), pp.97-121.
24. Hempel, C., 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.
25. Hempel, C. and Oppenheim, P., 1945. A Definition of „Degree of Confirmation“. *Philosophy of Science*, 12(2), pp.98-115.
26. Henderson, L., 2014. Bayesianism and Inference to the Best Explanation. *British Journal for the Philosophy of Science*, 65, pp.687-715.
27. Horwich, P., 1982. *Probability and Evidence*. Cambridge: Cambridge University Press.
28. Jeffrey, R., 1990. *The Logic of Decision* (2nd ed.). Chicago: University of Chicago Press.

29. Jeffrey, R., 1992. *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.
30. Kemeny, J. and Oppenheim, P., 1952. Degrees of Factual Support. *Philosophy of Science*, 19, pp.307-324.
31. Keynes, J., 1921/2004. *A Treatise on Probability*. New York: Dover.
32. Kolmogorov, A., 1933/1956. *Foundations of the Theory of Probability* (2nd ed.). New York: Chelsea Publishing Company.
33. Laudan, L. and Leplin, J., 1991. Empirical Equivalence and Underdetermination. *The Journal of Philosophy*, 88(9), pp. 449-472.
34. Lipton, P., 2001a. What Good is an Explanation? In: G. Hon and S. Rakover, eds. 2001. *Explanation: Theoretical Approaches and Applications*. Berlin: Springer. pp.43-59.
35. Lipton, P., 2001b. Is Explanation a Guide to Inference? A reply to Wesley C. Salmon. In: G. Hon, S. Rakover, eds. 2001. *Explanation: Theoretical Approaches and Applications*. Berlin: Springer. pp.93-120.
36. Lipton, P., 1991/2004. *Inference to the Best Explanation* (2nd Edition). London: Routledge.
37. Lombrozo, T., 2012. Explanation and Abductive Inference. In: *The Oxford Handbook of Thinking and Reasoning*, Keith J. Holyoak and Robert G. Morrison, eds. Oxford: Oxford University Press. pp.260-276.
38. Madigan, A.M. and McCourt, M., 2015. A New Inclination Instability Reshapes Keplerian Discs into Cones: Application to the Outer Solar System. *Monthly Notices of the Royal Astronomical Society*, 457, pp.89-93.
39. Nozick, R., 1981. *Philosophical Explanations*. Oxford: Clarendon Press.
40. Okasha, S., 2000. Van Fraassen's Critique of Inference to the Best Explanation. *Studies in History and Philosophy of Science*, 31, pp.691-710.
41. Popper, K., 1935/2005. *The Logic of Scientific Discovery*. London: Routledge.
42. Psillos, S., 2007. The Fine Structure of Inference to the Best Explanation. *Philosophy and Phenomenological Research*, 74(2), pp.441-448.
43. Quine, W.v.O., 1951. Main Trends in Recent Philosophy: Two Dogmas of Empiricism. *The Philosophical Review*, 60(1), pp.20-43.

44. Salmon, W., 2001a. Explanation and Confirmation: A Bayesian Critique of Inference to the Best Explanation. In: Hon and Rakover, eds. 2001. *Explanation: Theoretical Approaches and Applications*. Berlin: Springer. pp.61-91.
45. Salmon, W., 2001b. Reflections of a Bashful Bayesian: A Reply to Peter Lipton. In: Hon and Rakover, eds. 2001. *Explanation: Theoretical Approaches and Applications*. Berlin: Springer. pp.121-136.
46. Schupbach, J., 2011. Comparing Probabilistic Measures of Explanatory Power. *Philosophy of Science*, 78(5), pp.813-829.
47. Schupbach, J. and Sprenger, J., 2011. The Logic of Explanatory Power. *Philosophy of Science*, 78(1), pp.105-127.
48. Schurz, G., 2008. Patterns of Abduction. *Synthese*, 164, pp.201-234.
49. Schurz, G., 2014. *Philosophy of Science: A Unified Approach*. London: Routledge.
50. Scriven, M., 1959. Explanation and Prediction in Evolutionary Theory. *Science*, 130(3374), pp.477-482.
51. Talbott, W., 2001. Bayesian Epistemology. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Zalta, E.N., ed. [online] Available at: <https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/> [Accessed 30 January 2018].
52. Tentori, K., Crupi, V., Bonini, N. and Osherson D., 2007. Comparison of Confirmation Measures. *Cognition*, 103, pp.107-119.
53. Trujillo, C. and Sheppard, S., 2014. A Sedna-like Body with a Perihelion of 80 Astronomical Units. *Nature*, 507(7493), pp.471-474.
54. Van Fraassen, B., 1989. *Laws and Symmetry*. Oxford: Clarendon Press.
55. Weisberg, J., 2009. Locating IBE in the Bayesian Framework. *Synthese*, 167(1), pp.125-143.
56. Whewell, W., 1837. *History of the Inductive Sciences*. London: John W. Parker.