РЕЦЕНЗИЯ

от проф. д-р Лилия Александрова Гурова,
Нов български университет, професионално направление 2.3. Философия,
за дисертация на тема ***„Критични условия, при които се проявява ефектът на обърнатата базова честота“,***
представена от Йолина Атанасова Петрова за придобиване на образователната и научна степен „доктор“ в професионално направление 3.2. Психология.

Представеният текст е в обем от 132 с. и съдържа въведение, 11 глави, обща дискусия и заключение, описание на приносите, списък на използваната литература (120 заглавия) и 8 приложения.

1. Значимост на изследвания проблем

През 2015 г. Benishek и колеги обявяват, че близо 80% от диагностичните грешки в медицината се дължат на когнитивни предразсъдъци, сред които най-разпространени са свързаните с игнориране или неправилно интерпретиране на базовата честота (виж с. 15 от дисертацията). Ефектът на обърната базова честота, който се изследва в представения дисертационен труд, е един от най-интересните от тази група предразсъдъци (макар да се водят спорове дали е редно да бъде квалифициран като предразсъдък). Ефектът се проявява в това, че когато хората трябва да решат към коя от двете категории (*А* или *Б*), притежаващи обща характеристика, но и такава, която е уникална само за едната категория, трябва да отнесат обект, който притежава уникалните и за двете категории характеристики, но не притежава тяхната обща характеристика, те систематично причисляват двусмисления обект към по-рядко срещаната категория. Този ефект, описан преди 35 години от Медин и Еделсън (Medin &Edelson, 1988), оттогава насам е обект на оживени дискусии, като причината за това не са само нежеланите и потенциално опасни последствия от проявленията на ефекта в диагностичната практика. Ефектът на обърнатата базова честота представлява и сериозно теоретично предизвикателство, тъй като нито една от основните теории за категоризацията не предсказва, и съответно не обяснява, този ефект. Екземплярните модели например предсказват, че в ситуациите, в които при хората се наблюдава ефектът на обърнатата базова честота, те би трябвало да реагират точно обратното – т.е. да

причислят неясния обект към по-често срещаната категория, докато прототипните модели предсказват отсъствие на предпочитание, т.е. случайно отнасяне на неясния обект към едната или другата категория. Двете най-популярни, специално създадени обяснения на ефекта на обърнатата базова честота, базирани съответно на асоциативно учене и на изводи в съгласие с определени правила, също не получават еднозначна подкрепа. Тези предизвикателства са мотивирали Йолина Петрова да потърси отговори на някои от спорните въпроси, върху които е съсредоточена съвременната дискусия.

2. Цели и задачи на дисертационния труд, резултати

Представеното изследване си поставя три взаимосвързани цели (виж глава 4, с. 32): (1) да изследва ролята на ученето за появата на ефекта на обърната честота (*IBRE*); (2) да изследва основните алтернативни обяснения на този ефект; и по-специално, (3) да тества доминиращото към момента обяснение, представящо ефекта като продукт на асоциативно учене, водещо до формиране на асиметрични репрезентации на по-честата и на по-рядко появяващата се категория. За постигането на тези цели са внимателно планирани и проведени 6 експеримента и 1 симулация, приносът на които към реализацията на общите цели на дисертацията е описан по-долу.

Експеримент 1 (глава 5) възпроизвежда *IBRE* в класическата парадигма за учене чрез класификация, като са използвани прости визуални стимули, вместо традиционно предпочитаните вербални такива. Новото (освен стимулите), в сравнение с класическия експеримент на Крушке (1996), са вербалните протоколи, съдържащи дефинициите, които участниците в експеримента е трябвало да дадат на категориите, които са им били показвани. Вербалните протоколи ясно показват асиметрия в представянията на по-честите и по-редките категории: по-честите категории са преобладаващо дефинирани чрез техните две характеристики (общата за двете категории и уникалната за дадената категория), докато по-редките категории са дефинирани по-често от техните уникални характеристики. Йолина Петрова интерпретира този резултат като съвместим с обяснението на *IBRE* с формирането на асиметрични репрезентации в процеса на асоциативно учене на категориите.

Експеримент 2 (глава 6) е в голяма степен оригинален, тъй като за пръв път възпроизвежда *IBRE* в ситуация на учене чрез извод. Този тип учене потиска формирането на асиметрични репрезентация и затова експериментът е важен за разрешаването на спора дали асиметричните репрезентации са необходимо условие за наблюдаването на *IBRE*. Този експеримент показва, че асиметричните репрезентации не

са такова условие. Вербалните протоколи показват, че участниците наистина не са формирали асиметрични репрезентации и въпреки това демонстрират ефекта на обърнатата базова честота.

Експеримент 3 (глава 7) и експеримент 4 (глава 8) тестват възможно влияние на мотивацията върху ефекта на обърнатата базова честота, съответно въведена преди фазата на учене (експеримент 3) и преди фазата на тестване (експеримент 4). Тъй като не се наблюдава значима разлика в големината на ефекта при двата типа мотивация, но има такава спрямо ефекта, получен при експеримент 1 (при отсъствие на мотивация), Йолина Петрова заключава, че мотивацията не засяга само ученето, дори когато предхожда фазата на учене, следователно, усилването на *IBRE*, до което тя води, най-вероятно се дължи на някакъв тип рационални разсъждения.

Експеримент 5 (глава 9) има за цел да тества директно допускането, че *IBRE* се дължи на процеси, протичащи по време на фазата на учене, като при него тази фаза е максимално ограничена (всеки участник вижда най-много два пъти една и съща задача за категоризация). Макар и да не се наблюдава същинският *IBRE* (обръщане на предпочитанията за категоризация към по-рядката категория, когато е налице двусмислен обект, притежаващ и двете уникални характеристики на алтернативните категории, все пак в това условие участниците са значимо по-малко склонни да изберат по-честата категория, в сравнение с другите две двусмислени условия (когато е налице обект, притежаващ само общата характеристика и когато обектът притежава всички характеристики на двете категории).

Експеримент 6 (глава 10) въвежда контролно условие, при което двете категории от едната двойка категории се появяват с еднаква честота. Целта е да се провери дали наистина разликата в честотата на явяване на категориите е необходимо (Йолина го нарича „критично“) условие за генериране на *IBRE*. Експериментът показва, че разликата в честотата на появяване е необходимо/критично условие, доколкото в условието с изравнени честоти на двете категории, ефектът не се наблюдава.

В симулацията на *IBRE* с езиковия модел *GPT-3* (глава 11) моделът е поставен в ситуация, подобна на тази в класическата парадигма за установяване на *IBRE* при учене чрез класификация, с тази уговорка, че ученето е изключено, тъй като този модел не променя своите знания в резултат на решаване на поредица от задачи за категоризация. Оказва се, че въпреки отсъствието на учене в процеса на категоризация, езиковият модел *GPT-3*, подобно на хората, показва добре изразени предпочитания към категоризиране в по-рядката категория на неясния обект, притежаващ уникалните свойства на двете

алтернативни категории. Този резултат е особено любопитен, ако отчетем, че процесите на вземане на решение от страна на модели като *GPT-3* са много различни от онези, които протичат в ума на хората, поставени в подобни ситуации. Единствено можем с висока сигурност да твърдим (което Йолина Петрова прави), че научаването на категориите, което отсъства при *GPT-3*, не е необходимо условие за генерирането на ефекта на обърнатата базова честота.

3. Степен на познаване на състоянието на проблема и релевантната литература

Йолина Петрова демонстрира отлично познаване на изследванията, установяващи ефекта на обърнатата базова честота, както и на двата основни подхода към обяснението на ефекта, които поставят в основата му съответно асоциативното учене и изводите на базата на правила. Тя показва не само високо ниво на познаване (демонстрирано основно в глави 2 и 3), но и на разбиране на литературата, на която се позовава, което проличава в начина, по който са планирани изследванията, насочени към разрешаването на спорни въпроси в текущите дискусии, но най-вече в интерпретацията на получените резултати.

4. Съответствие на избраната методология на изследване на поставените цели и задачи на дисертационния труд

Йолина Петрова комбинира удачно класически експериментални изследвания, качествени изследвания (анализ на вербални протоколи) и симулации с езиков модел тип трансформатор, за постигане на основната цел, която си е поставила – да тества доминиращото обяснение на ефекта на обърната базова честота чрез асоциативно учене и на по-общото допускане, че процесите, водещи до ефекта на обърнатата базова честота са критично свързани с фазата на учене.

5. Наличие на собствен принос при събирането и анализа на емпиричните данни

Планираните от Йолина Петрова 6 експеримента са проведени от обучени от нея експериментатори. Получените сурови данни са обработени и анализирани от Йолина Петрова. Описаната в дисертацията компютърна симулация с използването на езиковия модел *GPT-3* е изцяло реализирана от Йолина Петрова.

6. Оценка на приносите на дисертационния труд

Самооценка на приносите на дисертационния труд се съдържа в самата дисертация (с. 105-106) и в автореферата към дисертацията. Своите приноси дисертантката е обобщила в три групи: методологични, емпирични и теоретични приноси. Без да омаловажавам методологичните и теоретичния принос, за мен най-интересни и важни с оглед на текущата дискусия са следните емпирични приноси: установява се, че (а) ефектът на обърнатата базова честота се наблюдава и в задачи за учене чрез извод, не само в задачи за учене чрез класификация; (б) формирането на асиметрични репрезентации не е необходимо условие за появата на този ефект; (в) наличието на фаза на учене също не е необходимо условие за генериране на ефекта. Тези експериментално установени находки (последната потвърдена също и от симулация с участието на *GPT-3*), освен че поставят под съмнение доминиращото обяснение на *IBRE* чрез асоциативно учене, разширяват представите ни за мащаба на този ефект и навеждат на мисълта, че е възможно ефектът на обърнатата базова честота да е продукт на различни механизми, които се проявяват с различна тежест в различните ситуации, подобно на самата категоризация, в контекста на която този ефект се проявява.

7. Оценка на публикациите по дисертационния труд

В автореферата са посочени 3 публикации по темата на дисертацията (в съавторство), като 2 от тях са в издания, реферирани в *SCOPUS*. Една от тези публикации (Petkov & Petrova, 2019) има 1 цитиране в списание, реферирано в *SCOPUS*, което не е автоцитат. Общо за периода от зачисляването в докторантура през 2018 г. досега, Йолина Петрова има 5 излезли от печат публикации (3 от тях в издания, реферирани в *SCOPUS*) и една (върху експеримент 1 и експеримент 2 от дисертацията), която предстои да бъде изпратена за публикуване в списанието *Memory & Cognition* (реферирано в *SCOPUS* и *Web of Science*). Броят и качеството на публикациите на Йолина Петрова свидетелстват за изградени качества на изследовател и умения за работа в интердисциплинарен екип.

8. Лични качества на кандидата

Познавам Йолина Петрова от времето, когато тя беше студент в магистърската програма по когнитивна наука в НБУ, но трайни впечатления имам от работата ѝ като редовен докторант към департамент „Когнитивна наука и психология“ в периода 2018 – 2020 г. Въпреки, че се наложи смяна на научния ръководител и темата на дисертацията

през втората година от докторантурата, Йолина успя да се организира и да събере необходимите кредити за отчисляване с право на защита предсрочно – за две години, вместо за предвидените в ЗРАСРБ 3 години. Освен със своята организираност, Йолина впечатлява и с мотивацията си да придобива нови знания и умения, които бързо усвоява до степен, която й позволява да преподава тези знания и умения на студенти. Не сме имали друг докторант, който без предходно математическо или техническо образование, да усвои определени методи за компютърно моделиране до степен, която да направи възможно използването им за решаване на изследователски задачи. Не на последно място, имам отлични впечатления от Йолина Петрова като преподавател. Още като докторант тя разработи и изнесе самостоятелно няколко лекции от курса „Понятия и категоризация“ в магистърската програма по когнитивна наука, като след назначаването й като асистент към департамента, пое този курс изцяло. Отзивите на студентите за начина, по който Йолина води този курс са отлични.

9. Мнения, препоръки и бележки

Общата ми оценка за представената дисертация е много висока. Получени са интересни резултати, които могат да допринесат съществено за развитието на дискусиите върху ефекта на обърнатата базова честота. Текстът е написан професионално, без излишно разводняване с цел трупане на страници, като в същото време е ясен и четивен. Още веднъж бих искала да обърна внимание на факта, че за постигането на поставените цели, докторантката е използвала умело различни методи – класически поведенчески експерименти, експлораторни анализи, анализи ва вербални протоколи, компютърни симулации, което се среща рядко в защитавани у нас дисертации в областта на психологията.

Въпросите и бележките ми са свързани главно с интерпретацията на основните резултати. Ако можем да приемем с висока степен на увереност, че тези резултати показват, че формирането на асиметрични репрезентации на представяните с различна честота категории не е необходимо условие за появата на ефекта на обърнатата базова честота, то заключението, че този ефект не е свързан с ученето се нуждае от по-внимателно прецизиране. Самата Йолина, както в дискусията върху резултатите от експеримент 5 (в който няма експлицитна фаза на учене), така и в общата дискусия (с. 100), допуска възможността за някакъв вид основано на екземпляри учене във фазата на тестване. Ако тя реши да публикува резултатите от експеримент 5 (което препоръчвам), евентуално съвместно с резултати от допълнителни изследвания, добре би било да се

помисли по какъв начин може да бъде потвърдена или отхвърлена хипотезата за наличието на латентно учене по време на тестването. Ако например се окаже, че големината на наблюдавания ефект на обърнатата базова честота зависи от продължителността на фазата на тестване (броят обекти, които участниците трябва да категоризират), то е много вероятно във формирането на този ефект да участва някакъв тип латентно учене.

Бих препоръчала на докторантката да продължи изследванията си със симулации на базата на езиков модел, защото този тип изследвания наистина позволяват прецизно изключване или манипулиране на факторите, които представляват интерес, както и да публикува резултатите от тези симулации, обръщайки внимание в дискусията на факта, че процесите на вземане на решение в езиковите модели са принципно различни от процесите, които предполагаме, че имат място при хората.

10. Заключение

Представеният дисертационен труд на тема „Критични условия, при които се проявява ефектът на обърнатата базова честота“ отговаря на всички изисквания, формулирани в ЗРАСРБ и Наредбата за развитие на академичния състав в НБУ за присъждане на образователната и научна степен „доктор“. Въз основа на това и на изтъкнатите по-горе достойнства на представената дисертация, ще гласувам „за“ присъждането на ОНС „доктор“ в професионално направление 3.2. Психология на Йолина Атанасова Петрова.

11.08.2023 г.

……..…….

..

/проф. д-р Лилия Гурова/