Министерство науки и высшего образования РФ
Российское психологическое общество
Ярославский государственный университет им. П. Г. Демидова
Лаборатория когнитивных исследований

Психология познания

Сборник материалов Всероссийской научной конференции памяти Дж. С. Брунера

Ярославский государственный университет
им. П. Г. Демидова

16–17 декабря 2022 г.

Ярославль
ЯрГУ
2023

УДК 159.9(063)
ББК 88.251я431
П86

Печатается в соответствии с решением оргкомитета
Всероссийской научной конференции «Психология познания»

Рецензенты:
В. Ф. Спиридонов – доктор психологических наук,
декан факультета психологии ИОН РАНХиГС, г. Москва
Е. С. Горбунова – кандидат психологических наук,
заведующая лабораторией когнитивной психологии пользователя
цифровых интерфейсов, НИУ ВШЭ, г. Москва

Ответственные редакторы:
И. Ю. Владимиров, С. Ю. Коровкин

П86 **Психология познания** : материалы Всероссийской научной конференции. ЯрГУ, 16–17 декабря 2022 г. /отв. ред. И. Ю. Владимиров, С. Ю. Коровкин – Ярославль : Филигрань, 2023. – 378 с. – ISBN 978-5-6049339-7-8.

В сборнике представлены материалы Всероссийской научной конференции «Психология познания», проходившей 16–17 декабря 2022 г. в ЯрГУ им П. Г. Демидова. Конференция посвящена памяти выдающегося психолога Дж. С. Брунера. В работе конференции приняли участие ученые ведущих исследовательских центров России по когнитивной психологии. Книга адресована специалистам в области когнитивной науки.

УДК 159.9(063)
ББК 88.251я431

ISBN 978-5-6049339-7-8

Одним из предикторов Ага является сложность ребуса: слишком простые задачи не провоцируют Ага, т. к. участники не успевают испытать непонимание. При этом оценки Ага выше для правильных ответов, чем для неправильных, что согласуется с данными аналогичных исследований (Salvi et al., 2016; Threadgold et al., 2018).

Список литературы

1. Морошкина Н. В., Аммалайнен А. В., Савина А. И. В погоне за инсайтом: современные подходы и методы измерения инсайта в когнитивной психологии // Психологические исследования. 2020. Т. 13. № 74. С. 5.

2. Salvi C., Constantini G., Bricolo E., Perugini M., Beeman M. Validation of Italian Rebus puzzles and compound remote associate problems // Behavior research methods. 2016. V. 48. N. 2. P. 664–685.

3. Sternberg R. J., Davidson J. E. The Nature of Insight. Cambridge, MA: The MIT Press. 1995.

4. Threadgold E., Marsh J. E., Ball L. J. Normative data for 84 UK English rebus puzzles // Frontiers in Psychology. 2018. V. 9. P. 2513.

УДК 159.9.07

Роль согласованности ответов при формировании доверия к советам от человека и искусственного интеллекта*

Л. А. Нездоймышапко [1,2]***, Р. В. Тихонов*** [2]
[1] *СПбГУ, Санкт-Петербург,*
[2] *НИУ «Высшая школа экономики», Санкт-Петербург*
e-mail: lanezdoymyshapko@edu.hse.ru

Аннотация. При принятии решений люди зачастую учитывают информацию, полученную из внешних источников – других людей или информационных систем. В нашем исследовании изучался процесс формирования доверия к советчику в задаче научения на основе нескольких признаков («multiple-cue learning»). Варьировались качество и источник советов (человек или «искусственный интеллект»). Было обнаружено, что участники в большей степени доверяли советам от «искусственного интеллекта», что проявилось как в объективных, так и в субъективных метриках доверия.

* Исследование выполнено за счет гранта Российского научного фонда (проект № 22-28-01456).

Ключевые слова: эпистемическое доверие, искусственный интеллект, согласованность ответов, научение на основе нескольких признаков.

В повседневных ситуациях мы нередко обращаемся к внешним источникам информации при вынесении суждений и принятии решений. Такими источниками могут выступать не только другие люди, но и информационные системы, которые с развитием технологий становятся все более точными и разнообразными. Исследователи показали (Kumar et al., 2021), что одних ситуациях люди проявляют недоверие к искусственному интеллекту («algorithm aversion»), в других же, напротив, слишком сильно придерживаются его рекомендаций («algorithm appreciation»). Однако лежащие в основе механизмы этих эффектов остаются неизвестными.

Известно, что учет мнений при вынесении собственных суждений происходит избирательно – какие-то советы действительно оказывают влияние на конечные решения, а какие-то игнорируются. Степень воспринимаемой надежности источника информации называют эпистемическим доверием (Echterhoff et al., 2005). Существуют разные факторы, оказывающие влияние на эпистемическое доверие. Одним из важнейших является качество информации, поступающей от него. Чем точнее советы, тем выше доверие к их источнику. Однако в отсутствие объективной обратной связи о качестве совета, люди могут оценивать источник только по косвенным признакам, таким как совпадение мнений в ситуациях с высокой степенью уверенности в собственном ответе. Влияние согласованности ответов с напарником на эпистемическое доверие было продемонстрировано в исследовании Крюгер (Krueger, 2017), где участники выносили оценочные суждения о личности другого человека. Однако неизвестно, как схожесть мнений влияет на формирование доверия в задачах, где ответы носят объективный характер (т. е. могут быть правильным или ошибочными).

Данное исследование призвано ответить на вопрос о том, как согласованность ответов влияет на доверие к напарнику. Предполагается, что качество и источник совета (искусственный интеллект или человек) повлияют на общее доверие к советчику и использование рекомендаций. Также предполагается, что степень согласованности мнений с советчиком будет положительно связана с доверием к нему.

Метод

Для проверки гипотез был организован онлайн-эксперимент по плану 2 (источник: искусственный интеллект или человек) x 2 (качество совета: хороший или плохой совет). Одна группа участников получала хороший совет от людей и плохой от ИИ, другая группа по-

лучала плохой совет от людей и хороший от ИИ. Источник совета отличался формулировкой. Для ИИ совет был сформулирован следующим образом: «Алгоритм искусственного интеллекта, который был обучен на этой задаче, предполагает, что правильный ответ – это ...». Для людей предложение выглядело так: «Человек, который ранее выполнил это задание, предполагает, что правильный ответ – ...». Все советы были основаны на алгоритмах линейной регрессии. В конце участники были проинформированы о том, что советник, который был описан как предыдущий участник, на самом деле был алгоритмом искусственного интеллекта. Качество совета было подсчитано как средняя ошибка совета. Хороший совет имел низкую среднюю ошибку (около 1 при максимальной ошибке 4), плохой совет имел высокую среднюю ошибку (более 3).

В качестве стимульного материала использовалась задача научения на основе нескольких признаков (multiple-cue learning), где люди обучаются предсказывать значение определенного параметра на основе набора нескольких других признаков, коррелирующих с итоговым значением. В данном эксперименте участникам предлагалась гипотетическая ситуация, в которой им нужно было представить себя на месте специалиста по подбору персонала, оценивающего кандидатов на некую должность. Они принимали решение о том, насколько кандидат подходит на работу по пятибалльной шкале (от «совершенно не подходит» до «полностью подходит») на основе результатов четырех тестов, каждый из которых мог принимать значения «очень низкий», «низкий», «высокий» и «очень высокий». Участникам заранее сообщалось о том, что тесты могут быть положительно или отрицательно связаны с успешностью кандидата, либо быть совершенно неинформативными. Однако оценить информативность теста они могли только косвенным образом на основе обратной связи.

Процедура состояла из обучающей серии (20 проб), тестовой серии (20 проб) и опросника. На этапе обучения участники выполняли задание, и получали обратную связь о правильности ответа. Этап обучения был организован таким образом, чтобы ответ (решение о том, насколько кандидат подходит для работы) имел нормальное распределение. Т. е. кандидаты были в основном средними, и лишь немногие из них были абсолютно подходящими или абсолютно неподходящими. В нашей задаче два теста в обучающей серии были положительно связаны с итоговой оценкой (корреляция более 0.8), а два выступали в качестве шума. На тестовой серии участникам необходимо было оценить кандидата и отметить свою уверенность в ответе. Затем они получали советы либо от искусственного интеллекта, либо от людей, хорошего или плохого качества. Участникам в первой половине экс-

перимента показывались рекомендации только от одного источника, затем из второго. Использовалась контрбалансировка порядка и качества источников. После получения совета участники повторно оценивали кандидата и собственную уверенность в ответе. Они также оценивали степень доверия к показанному совету от искусственного интеллекта или человека.

Всего в исследовании приняли участие 115 человек. Участники были набраны с помощью платформы Яндекс.Толока. Им было предложено принять участие в эксперименте за вознаграждение в размере 0,2$ за полное завершение. Они могли удвоить свое вознаграждение, если правильно ответили на вопрос для проверки внимания, а также получить бонус, если достигли высокой точности на этапе обучения. После оценки качества ответов часть участников была исключена. Окончательная выборка состояла из 96 человек.

Результаты

Средняя точность ответов в обучающей серии составила 78,9 % (SD = 4,7 %), что статистически значимо выше базового уровня успешности (70 %), которого можно добиться с помощью наиболее успешной эвристической стратегии (выбор среднего ответа по всех пробах), $t(95) = 18{,}54$, $p < 0{,}001$. Таким образом, участники продемонстрировали научение предъявленной закономерности.

Мы оценили влияние источника и качества советов на изменение ответа в сторону совета в ситуациях, когда мнения участников и советчиков различались с помощью логистической регрессии со смешанными эффектами по испытуемым. Зависимая переменная – сдвиг мнения в сторону совета. Независимые переменные – источник (человек или ИИ), качество (высокое/низкое), тестовая фаза, и взаимодействие факторов. Результаты показали, что люди реже прислушивались к советам от человека, чем от искусственного интеллекта (Odds Ratio (OR) = 0.24, 95 % CI [0.11; 0.56], p = 0.001), а также меньше опирались на советы во второй тестовой серии (OR = 0.14, 95 % CI [0.06; 0.32], p < 0.001). Кроме того, было обнаружено взаимодействие факторов (Источник * Тестовая фаза), свидетельствующее о более высоком влиянии мнений от советчиков-людей во второй тестовой серии (OR = 6.50, 95 % CI [2.01; 21.03], p = 0.002, по сравнению с первой. Качество советов не было статистически значимым предиктором сдвига мнений.

Для того, чтобы оценить роль совпадения исходных ответов с советчиком в формировании доверия к источнику информации, мы использовали линейную регрессионную модель со смешанными эффектами (по испытуемым) с воспринимаемым доверием в качестве зависимой переменной. Совпадение ответов, источник и их взаимодействие – предикторы. Обнаружено, что эпистемическое доверие к сове-

там от людей было ниже, чем к советам от ИИ (β = -0.52, 95 % CI [-0.69; -0.35], p < 0.001). Результаты также показали положительную взаимосвязь доверия с совпадением ответов (β = 1.44, 95 % CI [1.28; 1.59], p < 0.001) и взаимодействие факторов (Источник * Совпадение ответов), β = 0.41, 95 % CI [0.19; 0.62], p < 0.001, свидетельствующее о том, что совпадение ответов в большей степени влияло на формирование доверия к человеку, чем к искусственному интеллекту.

Наконец, мы включили в анализ фактор исходной уверенности, предположив, что именно уверенные совпадающие ответы будут влиять на формирование доверие к внешним источникам. Это предположение подтвердилось: вне зависимости от типа источника, уверенные совпадающие ответы выступали положительным предиктором воспринимаемого доверия (β = 0.92, 95 % CI [0.82; 1.02], p < 0.001).

Таким образом, мы обнаружили в задаче научения на основе нескольких признаков, что участники склонны опираться на советы от искусственного интеллекта, а также выше оценивают доверие к нему. Кроме того, было показано, что уверенность и совпадение ответов с советчиком также играют важную роль в формировании эпистемического доверия.

Список литературы

1. Echterhoff G., Higgins E. T., Groll S. Audience-tuning effects on memory: the role of shared reality // Journal of personality and social psychology. 2005. V. 89. N. 3. P. 257–276.

2. Krueger K. The impact of another person's responses to opinion communication: shared reality, epistemic trust, and belief certainty: diss. Kori Krueger. University of Pittsburgh, 2017. 66 p.

3. Kumar A., Patel T., Benjamin A. S., Steyvers M. Explaining Algorithm Aversion with Metacognitive Bandits // Proceedings of the Annual Meeting of the Cognitive Science Society. 2021. V. 43. №. 43.

Научное издание

Психология познания

Сборник материалов Всероссийской научной конференции
памяти Дж. С. Брунера
Ярославский государственный университет
им. П. Г. Демидова

16–17 декабря 2022 г.

Ответственные редакторы:
И. Ю. Владимиров, С. Ю. Коровкин

Верстка – Н. Ю. Лазарева, Н. Ю. Акатова

Подписано в печать 13.03.23. Формат 60x90 1/16.
Усл. печ. л. 11,00. Тираж 100 экз. Заказ № 23034.

Отпечатано в типографии ООО «Филигрань»
150049, г. Ярославль, ул. Свободы, д. 91.
pechataet@bk.ru