Vestnik YarGU. Seriya Gumanitarnye nauki. 2024. Vol. 18, No. 2
Journal homepage: http://j.uniyar.ac.ru/index.php/vyrgu

PSYCHOLOGY

Cognitive markers of the uncanny valley effect

V. D. Ilyushichev[1], A. V. Chistopolskaya[1], A. A. Kuritsyn[1]

[1]P. G. Demidov Yaroslavl State University, Yaroslavl, Russian Federation

DOI: 10.18255/1996-5648-2024-2-318-327

Research article
Full text in Russian

In his essay “The Uncanny Valley", Masahiro Mori notes the following: People will experience negative feelings towards a humanoid object if some of its features reveal its mechanical nature. When the nature of the object and its features are consistent, there is no such abrupt negative response. This transition to negative affect has been called the "uncanny valley" effect. One of the most popular explanations for this effect is the evolutionary hypothesis. Its main concept is the «pathogen avoidance system», an evolutionary mechanism responsible for the induction of aversion in response to the sight of a sick member of the species. Like the authors, we hypothesise that the same system is involved in the formation of the uncanny valley effect. In our study, subjects were asked to perform a task in which they had to match images of faces in three categories (healthy, infected, "uncanny") with one of two categories (alive/not alive). The result was that healthy faces were categorised with less accuracy and speed than uncanny faces. Our results provide partial support for the evolutionary hypothesis. In a subsequent experiment, we decided to clarify the influence of the atypicality of the faces presented on our earlier results by comparing the recognition success of three groups of faces: unusual/uncanny/typical. It was shown that unusual faces were recognised significantly more accurately than uncanny or typical faces.

Keywords: uncanny valley; categorization; evolutionary hypothesis; pathogen avoidance system; facial perception

INFORMATION ABOUT AUTHORS

Ilyushichev, Vladimir D.	E-mail: vladimirilyushichev@mail.ru
Chistopolskaya, Alexandra V.	E-mail: chistosasha@mail.ru ORCID iD: 0000-0002-6156-4876 Cand. Sc. (Psychology)
Kuritsyn, Alexander A.	E-mail: alexanderalkuritsyn@gmail.com

Funding: The President of the Russian Federation (project MK-495.2022.2)

Вестник ЯрГУ. Серия Гуманитарные науки. 2024. Том 18, № 2
веб-сайт: http://j.uniyar.ac.ru/index.php/vyrgu

Когнитивные маркеры эффекта «Зловещей Долины»

В. Д. Илюшичев[1], А. В. Чистопольская[1], А. А. Курицын[1]

[1]Ярославский государственный университет им. П. Г. Демидова, Ярославль, Российская Федерация

DOI: 10.18255/1996-5648-2024-2-318-327 Научная статья
УДК 159.93 Полный текст на русском языке

В эссе «Зловещая Долина» Масахиро Мори отмечает следующее: люди будут испытывать негативные переживания по отношению к человекоподобному объекту в том случае, если некоторые его черты выдают его механическую природу. В случае совпадения сущности объекта и его характеристик такого резкого негативного ответа не возникает. Этот переход к негативному аффекту получил название эффекта «Зловещей Долины». Одним из наиболее популярных объяснений этого эффекта стала эволюционная гипотеза. Основным её концептом является «система избегания патогенов» – эволюционный механизм, отвечающий за индукцию отвращения в ответ на вид больного представителя. Мы, как и авторы, предполагаем, что в формировании эффекта «Зловещей Долины» вовлечена та же система. В ходе нашего исследования испытуемым предлагалось выполнить задание на соотнесение изображений лиц трех категорий (Здоровые, Инфицированные, «Зловещие») к одной из двух категорий (живое/неживое). В результате было получено, что здоровые лица категоризируются с меньшей точностью и скоростью, чем «зловещие»; «зловещие» лица при этом не отличаются от инфицированных по показателям категоризации. Наши результаты частично подтверждают эволюционную гипотезу. В последующем эксперименте мы решили уточнить влияние нетипичности предъявляемых лиц на полученные нами ранее результаты. Для этого мы сравнили успешность узнавания трёх групп лиц: необычных/«зловещих»/типичных. Было показано, что необычные лица узнаются значимо более точно, чем «зловещие» или типичные лица.

Ключевые слова: Зловещая Долина; категоризация; эволюционная гипотеза; система избегания патогенов; восприятие лиц

ИНФОРМАЦИЯ ОБ АВТОРАХ

Илюшичев, Владимир Дмитриевич	E-mail: vladimirilyushichev@mail.ru Помощник исследователя, лаборатория когнитивных исследований
Чистопольская, Александра Валерьевна	E-mail: chistosasha@mail.ru; ORCID iD: 0000-0002-6156-4876 Кандидат психологических наук, лаборатория когнитивных исследований
Курицын, Александр Александрович	E-mail: alexanderalkuritsyn@gmail.com Помощник исследователя, лаборатория когнитивных исследований

Финансирование: Президент РФ (проект МК-495.2022.2).

Постановка проблемы исследования

В 1970 году профессор робототехники Токийского технологического института Масахиро Мори опубликовал эссе [1], в котором высказывалось предположение: отношение человека к человекоподобному роботу будет переходить от позитивной эмоциональной реакции и чувства эмпатии к ощущению страха и отвращения по мере приближения агента к реалистичному облику, но неполному его достижениию. Эта закономерность была представлена на графике, на котором ось абсцисс выражает уровень человекоподобности синтетического агента, а ось ординат – силу и модальность аффекта, вызываемого им.

Именно резкий переход к негативному аффекту был назван «Зловещей Долиной». В дальнейших работах, развивавших гипотезу Мори, достаточно абстрактное понятие негативного аффекта было конкретизировано до чувств, похожих на страх, тревогу, отвращение, потерю эмпатии и желание избежать агента [2–3]. Эссе вышло в малоизвестном японском журнале и долгое время не получало должного внимания, но с развитием робототехники и анимации концепт Мори стал более релевантным и начал использоваться довольно часто.

Эволюционная гипотеза возникновения эффекта «Зловещей Долины»

В процессе поисков объективных маркеров эффекта «Зловещей Долины» начинают появляться биологически ориентированные теории, наиболее распространённой из которых является эволюционная гипотеза [2]. В её свете роль механизма, обеспечивающего формирование эффекта, занимает система избегания патогенов. В обычной ситуации в ходе восприятия лица представителя собственного вида у человека актуализируется соответствующая категория. Эта категория содержит в себе ограниченный набор признаков. При восприятии лица больного представителя вида в поступающей информации о стимуле появляются признаки, выбивающиеся из категории вида (в случае с болезнью – это её различные проявления: сыпь, язвы и т. д.). В этом случае неоднозначность поступающей информации вызывает конфликт категоризации. В результате возникает негативная эмоциональная реакция. По тому же принципу описывается механизм «Зловещей Долины», только на место признаков болезни встают механические черты. При этом предполагается, что, если в возникновении эффекта «Зловещей Долины» действительно играет роль система избегания патогенов, то он должен встречаться не только у людей. Существует ряд работ, в которых предпринималась попытка изучения эффекта у приматов [2–4]. Что характерно для изучаемой проблемы, то эти работы также показывают противоположные результаты.

В работе Стекенфингера и Газанфара [2] проводилась регистрация движения глаз макак (m. fascicularis). В качестве стимульного материала использовались три типа изображений:

1. Реальные – фото макак того же вида;

2. Реалистичные – сгенерированные с помощью компьютерной графики лица макак того же вида с высокой степенью детализации;

3. Нереалистичные – компьютерно-сгенерированные лица макак того же вида с низкой степенью детализации.

Для операционализации переживания эффекта «Зловещей Долины» у приматов авторы опирались на работу Хамфри [5], согласно которой приматы склонны уделять меньшее зрительное внимание неприятным стимулам. Соответственно, если макака и вправду ощущает эффект «Зловещей Долины» во время просмотра изображения, она будет делать на нём меньшее количество фиксаций и общее время фиксаций будет ниже, чем в случае с другими стимулами. Авторы предполагали, что макаки будут испытывать эффект «Зловещей Долины» при восприятии реалистичного лица, в то же время нереалистичное лицо и реальное лицо не будут вызывать никакого специфичного отклика. Гипотеза авторов была подтверждена, макаки действительно уделяли меньшее внимание реалистичному лицу. Эти данные были интерпретированы как подтверждение эволюционной гипотезы. Однако в дальнейшем эта работа подвергалась существенной критике. Основной причиной для неё стало малое количество использованных стимулов, из-за чего различия между ними становились слишком существенными. Эта проблема создала пространство для потенциальных альтернативных объяснений. Например, разный уровень интереса может быть результатом новизны стимула, а не его «зловещести» [6].

В работе С. Б. Карп и коллег [4] представлена концептуальная репликация исследования Стекенфингера и Газанфара с расширенным варьированием реалистичности стимула (5 условий). Однако они не получили эффекта «Зловещей Долины» в своей выборке. При этом макаки уделяли большее внимание глазам во время восприятия реального изображения по сравнению с искусственными. Эти данные могут говорить о том, что восприятие реального и искусственного лица приматами различается, а эффект «Зловещей Долины» у животных не наблюдается.

Организация процедуры исследования

Участники исследования. В эксперименте приняли участие 48 студентов ЯрГУ им. П. Г. Демидова (45 женщин, 3 мужчины, средний возраст – 21.75 лет, SD = 5.5). Все испытуемые имели нормальное или скорректированное до нормального зрение.

Стимульный материал и процедура

Гипотеза: время и точность категоризации здоровых лиц будут значимо отличаться от лиц с инфекционными заболеваниями и «зловещих» лиц, их категоризация будет проходить дольше и с меньшей точностью. Время и точность категоризации лиц с инфекционными заболеваниями и «зловещих» лиц не будут иметь значимых различий.

Независимая переменная: группа лиц (здоровые лица, лица с инфекционными заболеваниями, «зловещие» лица).

Зависимая переменная: скорость и точность категоризации.

Процедура исследования: в качестве стимульного материала использовались изображения здоровых лиц из открытых баз данных; лиц с инфек-

ционными заболеваниями из медицинских справочников и искусственных («зловещих») лиц из открытых источников.

Стимульный материал делился на три группы изображений по 20 лиц в каждой: здоровые лица; лица с проявлениями инфекционных заболеваний; искусственные лица, вызывающие эффект «Зловещей Долины». Набор стимулов, вызывающих эффект «Зловещей Долины» был отобран по результатам предварительной экспертной оценки. Эксперты (40 человек, ср. возраст – 20.15 лет) оценивали каждый стимул при помощи 2 опросников, которые можно обнаружить в более ранних работах, посвященных «Зловещей Долине». Первый из них – опросник симпатии / нравищности все так Монатана (Monathants likeing question) – включает 4 полярные шкалы: Дружелюбный – Недружелюбный, Приятный – Неприятный, Добрый – Злой, Симпатичный – Несимпатичный [7]. Второй – опросник человекоподобности Пауэрса и Кильзера [8], включающий в себя следующие шкалы: Человекоподобный – Механический, Естественный – Искусственный, Осознающий себя – Не осознающий себя, Одушевлённый – Неодушевлённый. Стимулы, набравшие минимальные баллы по обоим опросникам, соответственно оцененные как наиболее искусственные и неприятные, вошли в основную серию эксперимента.

В основной серии испытуемым предлагалось выполнить задание на соотнесение изображений к одной из двух категорий (живое / неживое) при помощи клавиш влево и вправо. Эксперимент проводился в программе PsychoPy 2.0, стимулы предъявлялись на экране с частотой обновления 60 Гц. Время предъявления стимула составляло 1000 мс., стимулы разделялись фиксационным крестом, также предъявлявшимся на 1000 мс. Для адаптации к ритму предъявлений для всех испытуемых проводилась тренировочная серия, включавшая в себя 12 изображений, не входивших в основную серию.

Результаты

Для анализа скорости категоризации использовался U - критерий Манна-Уитни. Были получены значимые различия между группами здоровых и «зловещих» лиц ($U = 308$, $p < 0.01$) и между группами «зловещих» лиц и лиц с инфекционными заболеваниями ($U = 290$, $p < 0.05$). Различий в скорости категоризации для групп здоровых и инфицированных лиц обнаружено не было ($U = 221$, $p = 0.5831$) (рис. 1). Из анализа были исключены случаи, когда испытуемый не успевал отнести изображение к категории за отведенное время.

Для анализа точности категоризации использовался Хи-квадрат с поправкой Йейтса. При сравнении групп здоровых и инфицированных лиц не было обнаружено значимых различий ($\chi2 = 1.2$ $p = 0.27$), также не отличаются группы «зловещих» и инфицированных лиц ($\chi2 = 1.564$, $p = 0.2$). При этом значимые различия были зафиксированы при сравнении «зловещей» и здоровой групп ($\chi2 = 5.86$, $p < 0.05$) (рис. 2). Из анализа также были исключе-

ны случаи, когда испытуемый не успевал отнести изображение к категории за отведенное время.

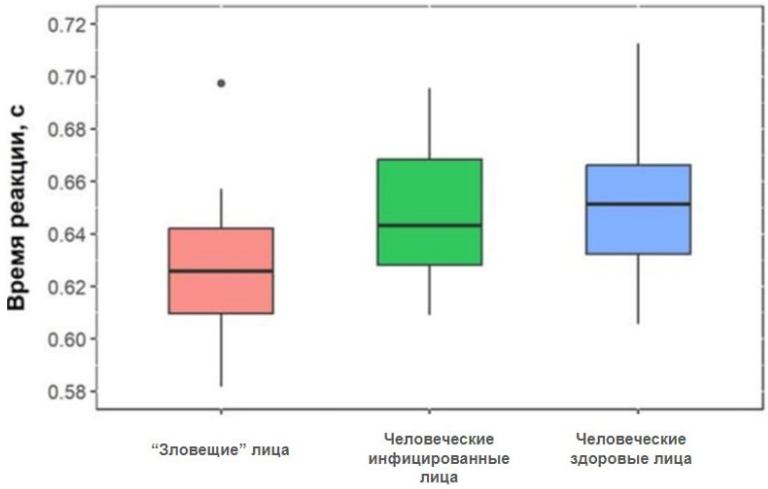

Рис. 1 Время реакции для «зловещих», инфицированных и здоровых лиц

Рис. 2 Показатели точности категоризации для «зловещих», инфицированных и здоровых лиц

Обсуждение

В результате анализа полученных данных можем сказать, что гипотеза частично подтверждена: по показателям точности категоризации «зловещие» лица действительно значимо отличаются от здоровых лиц, при этом различия с инфицированными лицами не были обнаружены. Однако отсутствие значимых различий между инфицированными и здоровыми лицами и более быстрая категоризация «зловещих» лиц ведет к появлению ряда вопро-

сов относительно эволюционной гипотезы «Зловещей Долины» и работы системы избегания патогенов [2] в целом. Во-первых, работа системы предполагает разрешение конфликта между категорией и выпадающими из нее признаками, вследствие чего мы предполагали, что при её активации категоризация будет проходить дольше, так как в существующих исследованиях уже было показано, что неконгруэнтные категории в стимуле увеличивают время реакции [9]. Однако мы получили фактически обратные результаты. В процессе анализа когнитивных систем мы использовали подход Дэвида Марра [10]. Согласно ему один из фундаментальных вопросов, который следует задавать любой когнитивной системе, состоит в том, какую информацию извлекает система.

Отвечая на него, можем сказать, что система избегания патогенов отвечает за детекцию черт, нехарактерных для представителя вида, эти черты присутствуют и в группе «зловещих» и в группе инфицированных лиц. В этом случае мы предполагаем, что более высокая скорость и точность категоризации «зловещих» лиц может быть объяснена большим числом и/или интенсивностью этих черт по сравнению с группой инфицированных. В то же время объяснение, вовлекающее систему избегания патогенов, в этом случае не может объяснить худшие результаты в группе здоровых лиц.

Одним из возможных объяснений может служить более фундаментальный механизм, ускоряющий обработку опасных стимулов. В этом случае наиболее важно как можно быстрее и точнее категоризировать стимулы из группы «зловещих» и инфицированных, чем из группы здоровых лиц [11]. Однако проверка этой идеи требует дальнейших исследований.

Другим объяснением может служить идея, использованная для операционализации эффекта «Зловещей Долины» в оригинальной статье: лица, активирующие систему избегания патогенов, удерживают на себе внимание на значительно меньшее время, чем здоровые лица [2].

Объяснение полученным данным может быть дано в рамках теории управления ошибками (Error management theory) [12]. Цена ошибок первого типа (ложноположительные) и ошибок второго типа (ложноотрицательные) не является одинаковой практически ни в одной из задач, которые встают перед человеком, поэтому мы склонны совершать наименее опасные ошибки в различных ситуациях. Мы предполагаем, что в случае со «зловещими» лицами более быстрая и точная категоризация по сравнению со здоровыми объясняется тем, что цена ошибки для отнесения к категории «неживое» в этом случае существенно меньше.

Также, возможно, мы не увидели ожидаемых результатов из-за слишком большого времени предъявления, за счёт чего, помимо работающей автоматически системы избегания патогенов, в процесс категоризации включались более высокоуровневые механизмы [11].

Эксперимент 2

Полученные в предыдущем эксперименте данные могут объясняться эволюционной гипотезой «Зловещей Долины», однако существуют и иные

варианты интерпретации. Наиболее очевидным из них является идея о влиянии не самого факта заболевания или «зловещести», а необычности лиц с их проявлениями, наличия у них ярких черт, с которыми человек не сталкивается в опыте. Для проверки этой гипотезы и уточнения отсутствия влияния фактора необычности на полученные результаты был проведен дополнительный эксперимент.

Участники исследования. В эксперименте приняли участие 40 студентов ЯрГУ им. П. Г. Демидова, они были поделены на две равные группы. Группа 1 (18 женщин, 2 мужчины, средний возраст – 19,85 лет, SD = 1,27), группа 2 (17 женщин, 3 мужчины, средний возраст – 20,3 года, SD = 1,81). Все испытуемые имели нормальное или скорректированное до нормального зрение.

Стимульный материал и процедура

Гипотеза 1: группы «зловещих» типичных и необычных лиц не будут значимо отличаться между собой по точности узнавания.

Гипотеза 2: эффект перевёрнутости лиц [13] будет более сильным для группы «зловещих» лиц, так как они воспринимаются как лица, однако их обработка требует больших затрат когнитивного ресурса вследствие включения системы избегания патогенов.

Независимая переменная: тип лица (необычное, «зловещее», типичное)

Зависимая переменная: точность узнавания

Процедура исследования: в качестве стимульного материала использовались изображения из открытых баз данных, в группе «зловещих» лиц для узнавания использовались изображения, отобранные в предыдущем исследовании, а также 20 «зловещих» лиц из открытых баз данных. Для каждой группы лиц использовалось по 40 изображений.

В основной серии испытуемым предлагалось выполнить задание на узнавание изображений. Эксперимент проводился в программе PsychoPy 2.0, стимулы предъявлялись на экране с частотой обновления 60 Гц. Эксперимент для каждой из групп состоял из трёх этапов стимулов: запоминания, перерыва, узнавания. На этапе запоминания время предъявления стимула составляло 3000 мс. и разделялось фиксационным крестом, также предъявлявшимся на 1000 мс. Затем следовал перерыв, длившийся 3 минуты. После перерыва проходила серия узнавания, стимулы предъявлялись на неограниченное количество времени, испытуемый сообщал, видел ли он стимул в предыдущей серии нажатием на клавишу.

Испытуемые были разделены на две группы по 20 человек. Участниками из первой группы задание на узнавание лиц выполнялось со стимулами в нормальной ориентации, для участников из второй группы стимулы были перевёрнуты на 180°.

Результаты

Для анализа точности узнавания использовался Хи-квадрат с поправкой Йейтса. Были обнаружены значимые различия в успешности узнавания необычных лиц: При сравнении с типичными лицами ($\chi^2 = 21.193$, $p < 0.001$), при сравнении с «зловещими» лицами ($\chi^2 = 36.398$, $p < 0.001$). Схожие данные

получены для узнавания и в инвертированной ориентации: при сравнении с типичными лицами ($\chi^2 = 14.061$, $p < 0.001$), при сравнении с «зловещими» лицами ($\chi^2 = 31.251$, $p < 0.001$). При этом типичные и «зловещие» лица по успешности узнавания не отличаются (рис. 3).

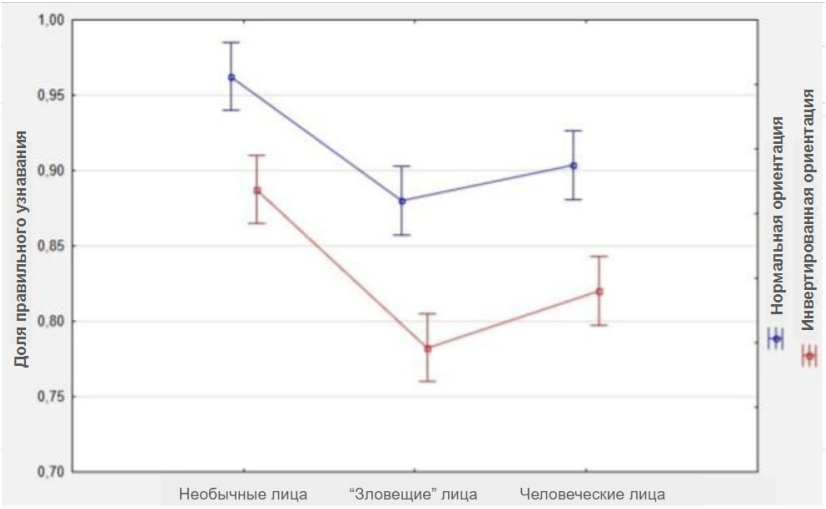

Рис. 3. Показатели точности узнавания лиц в зависимости от ориентации

Обсуждение

В результате проведённого эксперимента гипотеза 1 была частично подтверждена: группа необычных лиц значимо отличается от «зловещих» и типичных по точности узнавания. Мы считаем, что полученные данные говорят о большем количестве информации, которую предоставляют необычные лица; за счёт чего их узнавание оказывается более успешным. Соответственно, полученные нами в предыдущем эксперименте данные не могут объясняться только необычностью «зловещих» и инфицированных лиц. Гипотеза 2 не была подтверждена: при изменении ориентации стимулов «зловещие» лица продолжают узнаваться с точностью, неотличимой от нормальных. При этом необычные лица всё ещё узнаются значимо более успешно. Мы считаем, что это может быть объяснено наличием у них ярких особенностей, на которые испытуемый может опираться при узнавании. При этом выделяющиеся особенности наблюдаются также у «зловещих» лиц, но они при этом не сказываются на узнавании, что может говорить об отсутствии вклада «необычности» лица в возникновение эффекта «Зловещей Долины».

При сравнении результатов с прошлым экспериментом мы можем заметить, что «зловещие» лица категоризируются точнее здоровых, однако их узнавание происходит на том же уровне. Учитывая высокую долю правильных ответов в этих группах, мы можем предположить, что в наших данных наблюдается эффект потолка – задание оказалось слишком простым, чтобы обнаружить различия между группами. В дальнейших исследованиях для контроля этой побочной переменной мы можем создавать дефицит времени на узнавание.

Заключение

В ходе работы были получены данные, частично подтверждающие эволюционную гипотезу возникновения эффекта «Зловещей Долины». Однако стоит отметить, что они плохо согласуются с одной из базовых идей – конфликтом категоризации, предположительно возникающем при срабатывании системы избегания патогенов.

К важному ограничению исследования стоит отнести специфику «зловещих» лиц и их изначальный эмоциональный компонент. В нашей работе мы не уравнивали аффективную окраску в группах «зловещих», инфицированных, здоровых и необычных лиц, однако в дальнейшем это должно стать необходимой частью процедуры дизайна эксперимента.

Ссылки

1. Mori M., MacDorman K. F., Kageki N. The uncanny valley [from the field] // IEEE Robotics & automation magazine. 2012. Vol. 19, No. 2. P. 98–100.

2. Steckenfinger S. A., Ghazanfar A. A. Monkey visual behavior falls into the uncanny valley // Proceedings of the National Academy of Sciences. 2009. Vol. 106, No. 43. P. 18362–18366. DOI 10.1073/pnas.091006310.

3. Ho C. C., MacDorman K. F. Measuring the uncannyvalley effect // International Journal of Social Robotics. 2017. Vol. 9. P. 129–139. DOI 10.1007/s12369-016-0380-9.

4. Monkey visual attention does not fall into the uncanny valley / S. B. Carp [et all] // Scientific reports. 2022. Vol. 12, No. 1. Art. no. 11760. DOI 10.1038/s41598-022-14615-x

5. Humphrey N. K. Species and individuals in the perceptual world of monkeys // Perception. 1974. Vol. 3, No. 1. P. 105–114. DOI 10.1068/p030105.

6. Macaque gaze responses to the primatar: a virtual macaque head for social cognition research / V. A Wilson [et all] // Frontiers in Psychology. 2020. Vol. 11. P. 1645. DOI 10.3389/fpsyg.2020.01645

7. Monahan J. L. I don't know it but I like you: The influence of nonconscious affect on person perception // Human Communication Research. 1998. V. 24, No. 4. P. 480–500.

8. Powers A., Kiesler S. The advisor robot: tracing people's mental model from a robot's physical attributes. In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction. Association for Computing Machinery, New York, NY, USA. 2006. P. 218–225. DOI 10.1145/1121241.1121280.

9. Folstein J. R., Van Petten C., Rose S. A. Novelty and conflict in the categorization of complex stimuli // Psychophysiology. 2008. Vol. 45, No. 3. P. 467–479.

10. Marr D. Vision: A computational investigation into the human representation and processing of visual information. San Francisco. The MIT press. 2010. 445 p.

11. Del Giacco A. C., Ungerleider L. G., Yue X. Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization // Journal of Vision. 2018. Vol. 18, No. 12. P. 1–12. DOI 10.1167/18.12.3

12. Haselton M. G., Buss D. M. Error management theory: A new perspective on biases in cross-sex mind reading // Journal of Personality and Social Psychology. 2000. Vol. 78, No. 1. P. 81–91. DOI 10.1037/0022-3514.78.1.81

13. Yin R. K. Looking at upside-down faces // Journal of experimental pschology. 1969. Vol. 81, No. 1. C. 141.