

Правительство Российской Федерации
Федеральное Государственное Автономное Образовательное
Учреждение Высшего Образования
«Национальный Исследовательский Университет
“Высшая Школа Экономики”»

Факультет социальных наук Образовательная программа “Психология”
Направление подготовки 37.03.01 Психология

Курсовая работа на тему
«ЭФФЕКТ ЗЛОВЕЩЕЙ ДОЛИНЫ:
ИССЛЕДОВАТЕЛЬСКИЕ ПОДХОДЫ И МЕТОДЫ»

Выполнена **Коваленко Даниилом Сергеевичем**

Научный руководитель
Кандидат психологических наук, доцент, заведующая лабораторией
Когнитивной психологии пользования цифровых интерфейсов
Е. С. Горбунова

Москва 2020

Содержание

Введение	3
Глава 1. Теория и методология исследований феномена зловещей долины	8
Основные подходы	8
Методологические проблемы	11
Глава 2. Операционализация измерения антропоморфности	15
Использование реальных фотографий	15
Использование морфов	19
Использование реальных роботов	23
Размер континуума человекоподобия и количество экспериментальных условий	25
Глава 3. Операционализация эмоционального ответа	27
Измерение с помощью субъективной оценки	27
Измерение когнитивного конфликта	29
Альтернативные способы измерения	31
Глава 4. Направления развития исследований	35
Изучение индивидуальных и возрастных различий	35
Применение аппаратных методов	37
Математическое моделирование эффекта	38
Заключение	40
Список источников	41

Введение

С развитием компьютерной графики, технологий робототехники и дополненной реальности роботы и виртуальные персонажи в целом начинают играть всё большую роль в нашей жизни. Если раньше роботы использовались преимущественно для выполнения работы, подходящей под критерии “трёх ‘н’” - небезопасной, неинтересной, неприятной [Takayama et al., 2008], то сейчас они выполняют значительно более широкий спектр функций, достаточно важных как для отдельных людей, так и для общества в целом. Например, роботы могут использоваться во время ухода за пожилыми пациентами не только в качестве замены медперсонала, но и в целях снижения у этих групп пациентов чувства одиночества, а также их социализации и поддержания когнитивных функций [Birks et al., 2016]. Известно применение роботов в рамках терапевтических интервенций для детей с расстройствами аутистического спектра [Ishak et al., 2019], в рамках системы школьного образования для изучения технических предметов [Church et al., 2010], а также в более повседневных сферах, например, в качестве помощника посетителей аэропорта [Triebel et al., 2016]. Виртуальные персонажи, наиболее часто используемые в компьютерных играх, согласно данным последних исследований, играют немаловажную роль, будучи не только частью интерактивного окружения, но и агентами, воздействующими на игроков на эмоциональном уровне [Mousas et al., 2018; Felnhofer et al., 2018]. Однако сферы, в которых задействованы виртуальные персонажи не ограничиваются компьютерными играми. Известно их использование в рамках подготовки медперсонала для взаимодействия с пациентами [Volante et al., 2016], в качестве стимулов для исследований в области нейронауки [de Borst & de Gelder, 2015], “умных ассистентов” в программах с адаптивным интерфейсом и персонализированными планами обучения [Reategui et. al, 2008]. Таким образом, количество сфер применения роботов и виртуальных персонажей неуклонно возрастает, а вместе с тем возникает ряд вопросов, связанных с их дизайном и его ролью в восприятии данных агентов людьми, с ними взаимодействующими.

Одним из основных трудов, касающихся особенностей дизайна роботов, стала работа японского инженера М. Мори, опубликованная в 1970 [Mori, 1970] и переведённая на английский язык К. МакДорман в 2012 году [Mori et al., 2012]. В своей работе М. Мори предполагает, что вопреки нашим интуитивным представлениям, отношение между той степенью, в которой робот похож на человека и тем, насколько положительные эмоции он вызывает, не является линейным. М. Мори полагает, что по мере возрастания уровня человекоподобия робота наблюдается увеличение положительных эмоций, которые он вызывает, однако как только робот становится достаточно похожим на человека, но не идентичным ему, он может вызывать негативные эмоции (рис.1). Подобный “провал” в графике эмоционального ответа человека при взаимодействии с роботом и получил название эффекта зловещей долины (ЗД).

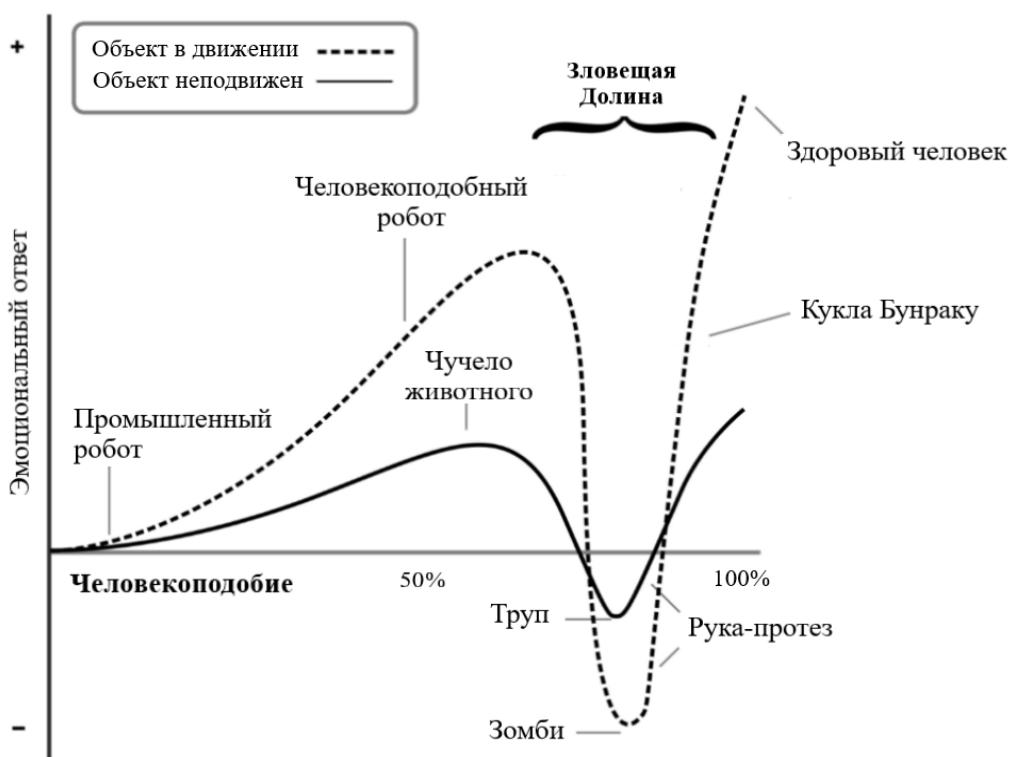


Рис. 1. Отношения между уровнем человекоподобия робота и валентностью эмоционального ответа, предложенные М. Мори.
Адаптировано из [MacDorman, 2005]

Несмотря на то, что гипотеза Мори была построена скорее на его личных наблюдениях и касалась непосредственно практических вопросов дизайна [Lay et al., 2016], в последние несколько десятилетий феномен зловещей долины привлек много внимания исследователей, так как вопрос о психологических механизмах формирования данного эффекта остался нераскрытым в оригинальной работе. В связи с этим эффект зловещей долины стал темой многих научных работ, ставивших своей целью ответить на два вопроса: существует ли подобный феномен вообще и какими могут быть механизмы его возникновения. На данный момент однозначного ответа на оба этих вопроса не существует. Как показывает последний мета-анализ работ в данной области [Kätsyri et al., 2015], эмпирических свидетельств, говорящих о воспроизводимости кривой, предложенной Мори, немного в сравнении с количеством посвящённых этому исследований, а однозначное объяснение феномена зловещей долины также отсутствует, хотя и существует ряд теорий, пытающихся его предложить. Большинство предлагаемых гипотез о природе ЗД берут свою основу в эмпирических исследованиях, и имеющиеся открытия в области достаточно неоднозначны, поскольку они завязаны на конкретной экспериментальной процедуре и тех манипуляциях, которые исследователи проводят исходя из своих предположений, также разнородных. В связи с этим мы предполагаем, что прогресс в области исследований феномена ЗД зависит как от совершенствования теоретических принципов объяснения эффекта, так и от развития используемых методических процедур.

В данной работе мы рассмотрим ряд теоретических взглядов, лежащих в основе исследований ЗД, и сделаем несколько предположений в отношении них. Однако их систематическое рассмотрение не является нашей главной задачей. Точно так же мы не ставим своей целью рассмотреть конкретные эмпирические данные в этой области. Вместо этого мы предлагаем рассмотреть те методы, которыми преимущественно пользуются исследователи, поскольку, по нашему мнению, их подробный анализ может как пролить свет на проблему

разнородности эмпирических свидетельств, так и помочь лучше понять теоретические основания многих работ.

На данный момент существует несколько крупных обзоров исследований в области феномена зловещей долины, однако систематический анализ методологии данного направления не проводился, хотя тема методов и затрагивалась в ряде работ. Так, в наиболее известном мета-обзоре Джари Катсири и его коллег, к которому мы ещё не раз обратимся, были рассмотрены вопросы операционализации независимой и зависимой переменной в рамках экспериментальных исследований ЗД [Kätsyri et al., 2015], хотя, как подчёркивали сами авторы, они охватили далеко не весь спектр возможных методов, так как фокусировались преимущественно на обзоре эмпирических свидетельств. Работой, посвящённой исключительно методам изучения зловещей долины стала статья Лэй и коллег, в которой был предложен список рекомендаций для исследователей [Lay et al., 2016], однако авторы исходили, во-первых, из перспективы “наивной гипотезы”, то есть возможности воспроизведения оригинальной кривой, предложенной Мори, и, во-вторых, из практических соображений применения конкретных процедур в исследовании, не уделив внимания теоретическим основаниям методов. Крупный обзор в области методологии исследований взаимодействия человек-робот (*human-robot interaction*, HRI) был проведён Е. Вайз и коллегами [Wiese et al., 2017]. Несмотря на то, что некоторые из их выводов применимы к области исследования феномена ЗД, работа касалось скорее применения методов нейронауки в изучении HRI, в связи с чем она не совсем релевантна исследуемой теме. Мы можем заметить, что при наличии огромного числа эмпирических работ в области феномена зловещей долины и всём многообразии использованных в них методов, комплексный и критический анализ методических процедур ещё не проводился, несмотря на его потенциальную значимость для развития всего направления.

В данной работе мы планируем закрыть имеющийся пробел в понимании методологии исследования зловещей долины. Необходимо отметить, что здесь и далее мы будем говорить об эффекте ЗД для статичных объектов, поскольку восприятие объектов в движении является отдельной обширной областью, развивающейся параллельно той, которой посвящена эта работа, и имеющей свои теоретические и методологические проблемы, которые мы не будем рассматривать. Таким образом, для данной работы мы выдвигаем следующие цели: рассмотреть наиболее известные методы и их теоретические обоснования, провести критический анализ возможностей их использования для решения основных методологических проблем феномена ЗД; предложить способы усовершенствования для уже имеющихся процедур и некоторые новые методы, которые можно было бы применять в рамках развивающихся направлений; рассмотреть перспективы развития области исследований зловещей долины в связи с развитием способов изучения данного феномена. Структура данной работы будет следующей: в главе 1 будут рассмотрены имеющиеся теоретические взгляды на эффект зловещей долины и основные методологические проблемы, связанные с изучением этого феномена; в главах 2 и 3 будет проведён анализ возможных решений выделенных методологических вопросов в перспективе различных теоретических подходов и основных используемых методов, будут рассмотрены преимущества и недостатки этих методов; в главе 4 будут обозначены основные направления исследований, которые смогут внести свой вклад в лучшее понимание феномена.

Глава 1. Теория и методология исследований феномена зловещей долины

Несмотря на то, что основной целью данной работы является критический анализ методов исследования феномена зловещей долины, этого невозможно сделать, не рассмотрев прежде те теоретические направления, которые лежат в основе использования тех или иных экспериментальных процедур. Иными словами, не зная возможных объяснений феномена, мы не можем определить, каким образом его лучше исследовать. В связи с этим в данном разделе мы обратимся к существующим объяснениям эффекта ЗД.

Основные подходы

Парадоксально, но первое возможное объяснение феномена появилось задолго до того, как сам феномен был сформулирован М. Мори. Речь идёт о работе немецкого психиатра Эрнста Йенча “О психологии зловещего” [Jentsch, 1906]. В ней предполагается, что чувство страха может быть вызвано неопределенностью, неуверенностью в том, является ли наблюдаемый человеком объект живым или нет. Несмотря на то, что во времена Йенча ещё не существовало роботов, а потому он использовал значительно более простые примеры, мы можем заметить, что сущность обозначаемого им явления достаточно хорошо согласуется с предложением Мори - негативные эмоции (в частности, чувство страха) возникает, когда мы не можем однозначно идентифицировать объект как живого человека или как известное нам живое существо. Сама по себе гипотеза Йенча не получила широкого распространения, возможно поскольку она сама требует эмпирической проверки, однако похожие идеи мы можем встретить и в более новых работах.

К другим ранним объяснениям эффекта зловещей долины могут относиться предложенные К. МакДорман эволюционная гипотеза избегания

патогенов [MacDorman et al., 2009] и гипотеза страха смерти [MacDorman, 2005a]. Согласно первому объяснению, наша неприязнь к похожим на людей роботам может появляться ввиду того, что они воспринимаются нами как люди с физическими дефектами, что вызывает соответствующую реакцию отвращения и, следовательно, избегающее поведение. Вторая гипотеза предполагает, что человекоподобные роботы напоминают нам мёртвые тела, то есть напоминают нам о смерти. Таким образом, страх, вызываемый роботами, по своей сущности является страхом смерти, а отвращение по отношению к ним является защитным механизмом, призванным снизить этот страх путём избегания объекта, его вызывающего. Данные гипотезы также практически не получили дальнейшего развития, что, вероятно, связано с методическими трудностями их проверки. Хотя как мы увидим далее, релевантные им методы всё-таки существуют.

Более современные объяснения феномена зловещей долины ищут причины негативного эмоционального ответа в когнитивных процессах восприятия и категоризации объектов. Одним из таких подходов является гипотеза перцептивного несоответствия, или рассогласования перцептивных признаков [Mäkäräinen et al., 2014; MacDorman & Chattpadhyay, 2016]. Согласно данному объяснению, негативный эмоциональный ответ при восприятии роботов возникает ввиду того, что во время кодировки отдельных признаков объекта эти признаки не согласуются друг с другом и, соответственно, с принятым паттерном восприятия. Например, негативные эмоции может вызывать персонаж с высоко реалистичными чертами лица, кожей, но имеющий “холодные”, сгенерированные компьютером глаза. Данная гипотеза, как показывают некоторые исследования, касается также и более целостного мультимодального восприятия - зловещими могут восприниматься агенты, у которых не согласуются уровни реализма внешнего вида и голоса [Mitchell et al., 2011]. Вариацией гипотезы перцептивного несоответствия можно считать гипотезу нетипичных признаков, согласно которой негативные эмоции могут быть вызваны при контакте с объектом, обладающим резко выделяющимися,

некоторыми для него чертами. В каком-то смысле это объяснение можно считать аналогичным предыдущему, однако, как указывают Катсири и коллеги, оно охватывает более широкий спектр явлений, не сводимых к пересечению признаков различных категорий (например, люди сискажёнными конечностями), а потому может быть выделено отдельно [Kätsyri et al., 2015].

Объяснение, которое во многом противостоит гипотезе перцептивного несоответствия, можно обозначить как теорию когнитивного конфликта, или категоризации. Сразу стоит уточнить, что единой теории когнитивного конфликта не существует. Содержательно в эту область попадает целый ряд исследований, отталкивающихся от предположения о том, что негативные эмоции, вызываемые антропоморфными роботами, происходят в качестве сопровождения [Weis & Wiese, 2017] или механизма подавления [Ferrey et al., 2015] того когнитивного диссонанса, который возникает при невозможности однозначно категоризировать тот или иной объект. Отдельно можно выделить гипотезу “избегания незнакомцев”, согласно которой объект, который сложно отнести к конкретной категории, попадает в новую, “незнакомую” категорию, которая рассматривается как потенциально опасная и, следовательно, вызывает негативные эмоции и поведение избегания [Yamada et al., 2013]. Можно заметить, что в некотором роде предположение о трудности категоризации и вызываемыми ею негативными эмоциями согласуется с гипотезой Йенча, однако, по нашим сведениям, такие параллели не проводятся на уровне самих эмпирических исследований. Более подробно конкретные предположения, основывающиеся на идеи когнитивного конфликта, будут рассмотрены в главе 2 при обсуждении использующихся методов, однако уже сейчас мы должны отметить, что данная концепция значительно более сложна для эмпирической проверки, чем гипотеза перцептивного несоответствия.

Новым подходом, появившимся в области исследований феномена ЗД, является объяснение с точки зрения социального функционализма [Olivera-La Rosa, 2018]. Согласно данному объяснению, негативные эмоции, вызываемые

слишком человекоподобными роботами, связаны в первую очередь с неясностью статуса робота как интенционального агента. Так, роботу могут быть приписаны психопатические черты, поскольку нередко его мимика невыразительна, или демонстрируемые им эмоции сложно интерпретировать; эта же проблема в целом затрудняет определение морального статуса робота, что может вызывать негативные эмоции исключительно нравственного содержания. Также автор гипотезы предполагает, что в возникновении негативной реакции на роботов может играть роль эвристика “красивый, значит хороший”, следствием которой выступает обратное утверждение - “некрасивый, значит плохой”. Данный подход, по нашему мнению, имеет ряд очевидных трудностей. Например, он не объясняет, почему именно и какие именно роботы воспринимаются некрасивыми и зловещими, а также не комментирует, возможно ли распространение данных объяснений на персонажей компьютерных игр, которые изначально выступают как интенциональные агенты и демонстрируют мимику. Тем не менее, интерпретация феномена зловещей долины с точки зрения социального функционализма может оказаться достаточно перспективной как для понимания эффекта в целом, так и для использования в сфере социальной робототехники. Поэтому в дальнейшем мы рассмотрим те методы исследования, которые были предложены автором, и выскажем предположения о том, какие ещё методы могли бы быть использованы в этом подходе.

Методологические проблемы

До этого момента мы говорили об эффекте зловещей долины в терминах схожести агента с человеком и соответствующей ей валентности эмоционального ответа при восприятии этого агента, однако данные термины могут показаться слишком общими, и они практически не используются в исследованиях. Мы же использовали их во избежание путаницы, связанной с операционализацией гипотезы М. Мори. Мы считаем необходимым

подчеркнуть, что эффект зловещей долины является отношением между двумя показателями, единой операционализации которых не существует до сих пор, в чём и заключается одна из крупнейших трудностей исследований в области.

Так, степень сходства с человеком, иными словами человекоподобие или антропоморфность (именно эти термины мы будем использовать в дальнейшем по аналогии с наиболее часто употребляемым в исследованиях “human-likeness”), не определены единым образом. Несмотря на то, что, как правило, используется подобие человеку относительно робота или компьютерной модели, мы не можем наверняка утверждать, что это два равнозначных вида антропоморфности. Аналогичным образом, мы не можем говорить, что эффект, наблюдаемый на континууме человек-животное, аналогичен эффекту, который можно зафиксировать на континууме человек-робот. Наиболее чётко это разделение было выражено в работе Н. Хаслама, посвящённой процессу дегуманизации [Haslam, 2006]. Автор утверждает, что существует два вида дегуманизации: первый связан с чертами человека как живого существа (измерение “человеческой природы”), то есть наличием базовых эмоций, способности к выражению теплоты и т.д.; второй вид связан с чертами человека как исключительного и единственного в своём роде представителя животного мира (измерение “человеческой уникальности”), то есть с способностью к социализации и мышлению. Уже на основании этого мы можем предположить, что континуумы человек-робот и человек-животное, в целом подходящие под измерение человекоподобия, по своему содержанию могут быть разными, как могут гипотетически отличаться различные измерения антропоморфности внутри каждого из них.

Ещё один вопрос, связанный с операционализацией человекоподобия, заключается в размере того континуума, который мы используем. Так, на рис. 1 показано, что, согласно оригинальной гипотезе М. Мори объектом, относительно которого начинается континуум антропоморфности - промышленный робот, то есть объект, который имеет минимум черт, схожих с человеком. Хотя, как уже

указывалось в предыдущих работах в области, примеры, используемые Мори, скорее анекдотичны [Kätsyri et al., 2015] и не могут быть использованы в исследованиях, это не решает вопрос размера континуума. Так, в исследовании К. МакДорман и Д. Чаттопадьи был использован набор стимулов, варьирующихся от компьютерной модели персонажа до реального человека. Согласно результатам, наиболее негативный эмоциональный ответ вызвала именно компьютерная модель, в то время как в дальнейшем валентность ответа увеличивалась линейно [MacDorman & Chattopadhyay, 2016]. Позже это исследование подверглось критике за использование такого ограниченного набора стимулов [Kawabe et al., 2017], поскольку компьютерная модель как раз и является объектом, “проваливающимся в зловещую долину”. Этот пример, по нашему мнению, иллюстрирует важность корректной операционализации спектра человекоподобия, без которой невозможно получить валидные результаты. В главе 2 мы более подробно коснёмся этого вопроса, обсуждая подбор стимульных характеристик.

Вторая важная методологическая проблема исследований зловещей долины, получившая значительно больше внимания исследователей (например: Lay et al., 2016; Wang & Rochat, 2017; Olivera-La Rosa, 2018) - операционализация собственно эмоционального ответа, выступающего зависимой переменной в исследованиях. В своей оригинальной работе М. Мори использовал термины “shinwakan” и “bukimi” для обозначения положительного и отрицательного эмоционального ответа соответственно. Оба этих слова являются неологизмами даже на японском языке и обладают специфическими значениями, точных аналогов которым нет в английском [Ho & MacDorman, 2010] и, вероятно, других языках, что делает их использование в исследованиях крайне сложным. В следующей главе мы обсудим предложенные исследователями способы перевода и операционализации этих понятий, однако уже сейчас стоит отметить, что эмоции, которые вызываются агентами на континууме человекоподобия, должны не просто обладать валентностью. Они также должны быть наделены

специфическим содержанием: под положительными эмоциями мы можем подразумевать радость, интерес, чувство “близости”, а под негативными - страх, отвращение, презрение и так далее.

Последний, но не по значимости, вопрос, который как правило игнорируется большинством исследователей, но которому мы считаем нужным уделить внимание - общая характеристика кривой “зловещей долины”. В своём обзоре Катсири и коллеги указывали на то, что большинство современных исследований фокусируется исключительно на “отрицательной” части феномена, а именно на возникновении негативного эмоционального ответа при восприятии слишком человекоподобных роботов [Kätsyri et al., 2015]. Однако оригинальная гипотеза подразумевает также, что по мере увеличения человекоподобия перед “провалом” будет наблюдаться пик в эмоциональном ответе на роботов, которые ещё недостаточно антропоморфны, чтобы “упасть в зловещую долину”. Мы считаем, что эта часть гипотезы не должна оставаться незамеченной, а потому в дальнейшем уделим внимание также и возможностям её эмпирической проверки.

Глава 2. Операционализация измерения антропоморфности

В предыдущей главе были определены основные теоретические направления, в рамках которых происходят исследования феномена зловещей долины, а также выделили ряд методологических трудностей в понимании самого феномена. В данной главе мы перейдём к рассмотрению подходов к решению первой из обозначенных нами проблем, то есть операционализации континуума человекоподобия, с точки зрения различных направлений исследований. Мы рассмотрим такие способы операционализации, как использование реальных фотографий, морфов и реальных роботов, а также коснёмся проблемы размера континуума антропоморфности.

Использование реальных фотографий

На уровне дизайна эксперимента определение и операционализация измерения человекоподобия заключается в проблеме создания ряда стимульных материалов, которые отражали бы достаточно плавный переход между категориями робота (или другого существа) и человека.

Самым простым и наиболее доступным способом создания такого континуума является использование реальных фотографий людей и роботов с их ранжированием по схожести с человеком. Данный способ подбора стимульных материалов, обладает, несомненно, рядом преимуществ, помимо простоты их поиска. Так, применение реальных фотографий позволяет наиболее точно воспроизвести оригинальный континуум человекоподобия с соответствующими примерами, приблизившись к проверке гипотезы М. Мори, а также обладает большей экологической валидностью. В то же самое время при их использовании в качестве стимулов возникает ряд трудностей.

Трудности технического характера связаны с подбором достаточно гомогенных фотографий людей и роботов таким образом, чтобы избежать смешений, связанных с ракурсом фотографии, воспринимаемыми чертами и

выражениями лица, позой тела (в случае фотографий агентов в полный рост). Данные проблемы решаются исследователями путём составления определённых критериев для поиска и включения фотографий в стимульный ряд. Так, С. Вонг и П. Рошат в своём исследовании уравнивают между стимулами распределение характеристик пола, относительного возраста, расы, и выделяющихся черт лица, таких как растительность [Wang & Rochat, 2017]. М. Матур и Д. Ричлинг в своей работе предлагают более строгие и подробные критерии: робот или человек должен быть сфотографирован фронтально, все его черты должны быть видимы; робот должен быть изначально предназначен для социального взаимодействия с человеком; робот не должен представлять собой известного персонажа, быть распространённой игрушкой; на фотографии не должны присутствовать иные части тела, кроме лица [Mathur & Reichling, 2016]. В ряде других работ также предлагают использовать обрезанные фотографии лиц определённого формата во избежание смешений, связанных с наличием фона (например: Wang & Rochat, 2017; Weis & Wiese, 2017), хотя это требование более актуально для исследований, где в качестве стимулов используются морфы, о которых мы поговорим далее. В более поздней работе М. Матур и соавторов критерии были дополнены тем, что робот не должен демонстрировать определённых эмоций [Mathur et al., 2020]. Это требование мы считаем крайне важным, поскольку выражение роботом эмоций может вызывать соответствующую реакцию испытуемых, влияя на их эмоциональный ответ и способствуя, таким образом, смешению эффектов внешности робота и демонстрируемых выражений “лица”. Использование фотографий роботов и людей в полный рост распространено в меньшей степени и, стоит отметить, критериев отбора подобных стимульного материала на данный момент предложено не было. Тем не менее, мы можем предположить, что в случае использования таких стимулов фотографии должны представлять собой снимки агентов в полный рост или по пояс, в нейтральной позе, то есть с опущенными руками. Отдельным вопросом является наличие одежды, так как, с одной стороны, она является неотъемлемой характеристикой

людей, а с другой стороны не относится напрямую к антропометрическим параметрам. Иными словами, одежда может являться признаком “человеческой уникальности”, не относясь к “человеческой природе” в терминологии Н. Хаслама, то есть её наличие в ряде стимулов может содержательно определять используемый континуум. Так или иначе, полноценные фотографии роботов и людей являются более сложными стимулами, и требования к их подбору ещё должны быть разработаны.

Другим, значительно менее простым вопросом, связанным с использованием реальных фотографий, является определение положения каждой из них в спектре человекоподобия, то есть оценка того, в какой степени тот или иной агент похож на “эталонных” робота или человека. Вполне естественно, что, как правило, данная проблема решается путём субъективной оценки, производимой либо экспериментаторами, либо испытуемыми. Второй вариант, по замечанию Лэй и соавторов, является довольно спорным, поскольку если при предъявлении стимула испытуемых просят оценить и то, в какой мере он похож на человека, и то, насколько положительные эмоции он вызывает, могут возникнуть связанные с этим систематические смешения [Lay et al., 2016]. Поэтому при подборе стимулов необходимо распределять их на континууме “человекоподобия” исходя из оценок независимой группы экспертов в рамках пилотного исследования, как это было сделано в исследовании М. Матур и соавторов. Экспертов просили отметить, насколько каждый из представленных агентов является человекоподобным или механическим по шкале от -100 (“совершенно механический”) до +100 (“совершенно подобный человеку”). Аналогичным образом было проконтролировано отсутствие у агентов эмоциональных выражений лица [Mathur et al., 2020]. Помимо ранжирования представленных стимулов с помощью шкалирования также можно использовать задания с вынужденным двухальтернативным выбором, по итогам которого стимулы распределяются по выраженности своего сходства с человеком.

Описанные выше трудности определения степени “человекоподобия” агентов более характерны для исследований, в которых используется большой спектр стимулов, однако более классическим для работ в области феномена зловещей долины является предъявление трёх-пяти стимулов, соответствующих различным экспериментальным условиям. В таком случае задача подбора стимулов упрощается. Например, в исследовании с тремя условиями достаточным является подбор фотографий, соответствующих условиям “робот”, “антропоморфный робот” и “человек”. Однако и в таком случае с помощью пилотного исследования стоит удостовериться, что выбранные стимулы занимают ожидаемое положение в спектре антропоморфности. Более подробно мы рассмотрим этот вопрос, говоря о размере используемого континуума.

Сейчас же мы считаем нужным отметить, что реальные фотографии могут быть использованы в исследованиях, отталкивающихся как от теории перцептивного несоответствия, так и от гипотезы когнитивного конфликта. Примечательным в данном случае является исследование М. Страйт и коллег, в котором использовались фотографии, соответствующие шести условиям - робот, человек, робот и человек с нетипичными чертами, робот и человек неопределенной категории [Strait et al., 2017]. Таким образом, авторы пытались решить вопрос о том, какое из двух конкурирующих направлений лучше объясняет эффект зловещей долины, и, согласно полученным результатам, оба подхода предсказывают возникновение феномена. Мы полагаем, что к данным результатам стоит относиться скептически, поскольку при подборе стимульных материалов авторами был допущен ряд неточностей. Несмотря на то, что они ориентировались на ранее упомянутые критерии М. Матур, в их работе не были проконтролированы половозрастные характеристики запечатлённых на фотографиях агентов, а также не были разделены фотографии только лиц или агентов целиком. Способы, которыми авторы определяли то, относится ли объект к неопределенной категории или обладает нетипичными для своей категории чертами, были также достаточно субъективными. Несмотря на эти

недочёты, мы можем заметить, что использование реальных фотографий может быть перспективным с точки зрения как гипотезы перцептивного несоответствия, так и гипотезы когнитивного конфликта - разумеется, при соблюдении необходимых требований для выбора стимулов.

Использование морфов

Способом подбора стимульных материалов, обычно противопоставляемым использованию реальных фотографий, является использование морфов. Стоит сразу отметить, что методология морфов включает применение и фотoreалистичных морфов, то есть с использованием перехода между двумя фотографиями, и морфов, полученных путём поэтапной трансформации 3D-моделей. Мы рассмотрим оба этих метода создания стимульных материалов.

Впервые фотoreалистичные морфы были использованы в исследовании Ю. Ямада и коллег, предложивших в качестве объяснения феномена зловещей долины гипотезу “избегания незнакомцев” (categorization-based stranger avoidance) [Yamada et al., 2013], о которой было сказано в главе 1. Таким образом, с самого начала своего существования методология морфов применялась преимущественно в парадигме категоризации. Среди более поздних исследований данный метод использовался в работах П. П. Вайс и Е. Вайз [Weis & Wiese, 2017; Wiese & Weis, 2020], в которых когнитивный конфликт рассматривался как непосредственный источник негативных эмоций, возникающих при сложности категоризации неоднозначных агентов.

Несомненным и, вероятно, основным преимуществом применения фотoreалистичных морфов является возможность чётко проконтролировать переход от категории “робот” к категории “человек”, при этом установив фиксированное изменение степени человекоподобия между двумя соседствующими стимулами, на что указывали М. Читам и Л. Джанке в своём фМРТ-исследовании [Cheetham & Jancke, 2013]. Данное преимущество также

позволяет варьировать количество стимулов в ряду, изменяя “шаг” перехода между ними. Кроме того, фотогенеративные морфы позволяют моделировать различные континуумы человекоподобия, то есть переходы как от животного или робота к человеку, так и, например, от мягкой игрушки к человеку [Wiese & Weis, 2020].

Тем не менее, недостатком фотогенеративных морфов является наличие у некоторых стимулов, занимающих положение в середине ряда, накладывающихся друг на друга черт различных агентов [Kätsyri et al., 2015]. Так, в стимулах, использованных в работе Ю. Ямада, чёлка персонажа мультфильма очевидным образом накладывается на лицо реального человека, что, несомненно, снижает реалистичность агента (рис. 2). Недостаточная реалистичность стимулов является ещё одним направлением критики методологии морфов. Так, М. Матур и коллеги указывают на то, что агенты, сгенерированные путём перехода между категориями, являются априори нереалистичными, поскольку они не встречаются в реальной жизни, а потому не могут быть использованы в исследованиях феномена зловещей долины [Mathur et al., 2020]. Мы также считаем нужным отметить, что использование морфов не позволяет охватить весь предложенный М. Мори континуум, поскольку переход между индустриальным роботом, не обладающим человеческими чертами, и фотографией человека, невозможно создать таким образом - в противном случае используемые стимулы будут обладать максимальным количеством нерелевантных перекрывающихся признаков или представлять собой полупрозрачное наложение фотографий. Таким образом, рассмотрение эффекта ЗД в его “наивном” варианте, предполагающем в том числе использование примеров М. Мори, конечно, исключает применение фотогенеративных морфов. Но это не означает, что в исследованиях когнитивного конфликта как источника возникновения эффекта данный метод не может быть применён, хотя его экологическая валидность остаётся спорной.



Рис. 2. Примеры стимульных материалов из работы Ю. Ямада и коллег [Yamada et al., 2013]

Альтернативой фотогенеральным морфам является применение сгенерированных 3D-моделей создающих переход от робота или другого объекта к человеку. Такие модели использовались, например, в исследованиях Е. Бродбента и коллег [Broadbent et al., 2013], А. Ферри и коллег [Ferrey et al., 2015]. Данный метод создания стимулов обладает рядом преимуществ, свойственных методу морфов в целом, то есть позволяет проконтролировать степень изменения человекоподобия и создать достаточно продолжительный континуум. Применение 3D-моделей решает проблему наложения нерелевантных черт, а также, по нашему мнению, может решить вопрос ограниченности континуума морфов, поскольку с помощью техник моделирования можно создать плавные переходы от индустриального робота к человекоподобному и далее к человеку - хотя, по нашим сведениям, попыток сделать это в исследованиях феномена 3Д ещё не предпринималось. Кроме того, 3D-морфы могут обладать большей экологической валидностью, так как 3D-модели могут встречаться в компьютерных играх, мультипликации и фильмах.

Тем не менее, основной проблемой данного метода является, с другой стороны, недостаточная реалистичность используемых моделей даже относительно применяемых в компьютерных играх и кинематографе. Ни в одном из известных нам исследований не использовались фотогенеральные 3D-модели с достаточно проработанными деталями. Это означает, что крайние точки континуума, которые должны репрезентировать робота (или животное) и здорового человека, смешены относительно тех положений на спектре

человекоподобия, которые должны реально занимать. Так, недостаточно реалистичная 3D-модель человека может быть “зловещей” сама по себе, вне зависимости от того, какое положение она занимает. Это было показано в исследовании К. МакДорман и Д. Чаттопадьи [MacDorman & Chattopadhyay, 2016], где из всех представленных на континууме стимулов наиболее негативные эмоции вызывала как раз компьютерная модель, а не какой-либо агент, занимающий промежуточное положение между ней и человеком. В исследовании Ферри и соавторов такой результат не был получен, и 3D-модель человека, находившаяся на положительном полюсе антропоморфности, была оценена наиболее позитивно [Ferrey et al., 2015]. Мы можем предположить, что это произошло, поскольку данная модель была в принципе недостаточно реалистичной, чтобы “провалиться в зловещую долину”, или была оценена положительно относительно других стимулов, в действительности менее “приятных”.

Метод 3D-моделирования стимулов, в отличие от использования фотореалистичных морфов, не связан с каким-либо конкретным направлением исследований. Он применяется сторонниками теории перцептивного несоответствия [MacDorman & Chattopadhyay, 2016], поскольку создание 3D-морфов позволяет осуществлять отрисовку и проработку одних черт лица агента отдельно от других, таким образом манипулируя их соответствием друг другу. Сторонники теории когнитивного конфликта также активно используют технологию 3D-моделирования по уже упомянутым нами причинам. Отдельно стоит отметить, что использование 3D-морфов позволяет применять в исследованиях 3Д искусственно созданные категории, манипулируя степенью их “знакомости”. Так, Т. Бурлей и Дж. Шонер [Burleigh & Schoenherr, 2015] в своём эксперименте создали у испытуемых категории “Cax” и “Miv”, соответствующие таким агентам, как “инопланетянин”, “рептилия” и “зверь”, в дальнейшем проверяя эмоциональную реакцию, возникающую при появлении агентов, принадлежащих к неоднозначной категории.

Таким образом, применение 3D-морфов является, по нашему мнению, наиболее перспективным методом создания стимульных материалов ввиду своей универсальности. Однако он также требует использования продвинутых, фотореалистичных моделей во избежание снижения экологической валидности и смещения стимулов с ожидаемого положения на континууме антропоморфности.

Использование реальных роботов

Довольно перспективным способом исследования феномена ЗД является использование в качестве стимулов реальных роботов, взаимодействующих с людьми (например: Złotowski et al., 2015; Koschate et al., 2016). Такой метод операционализации человекоподобия несомненно, не подойдёт для исследования конкретных перцептивных признаков, влияющих на возникновение эффекта, однако может применяться в исследовании “мультимодального” феномена зловещей долины, а также для проверки устойчивости эффекта при взаимодействии человека и робота. Тем не менее, наибольшая сложность, связанная с использованием реальных роботов, состоит в контроле смещений, вызванных большим количеством внешних различий агентов, что делает их практически невозможными для сопоставления. Например, в исследовании Я. Злотовски и коллег были задействованы две модели роботов: Geminoid-2, высокого уровня антропоморфности, “проваливающийся” в зловещую долину, и Robovie R2, низкого уровня схожести с человеком и вызывающий нейтральный ответ (рис. 3) [Złotowski et al., 2015]. Данные роботы, по нашему мнению, действительно плохо подлежат сравнению, и определение их положения на континууме антропоморфности является достаточно умозрительным.

Эта проблема была решена в работе М. Пейтсел и Дж. Кастеллано, где для взаимодействия с испытуемыми использовались не реальные роботы, а полупрозрачная голова робота, на лицевую поверхность которой была наложена

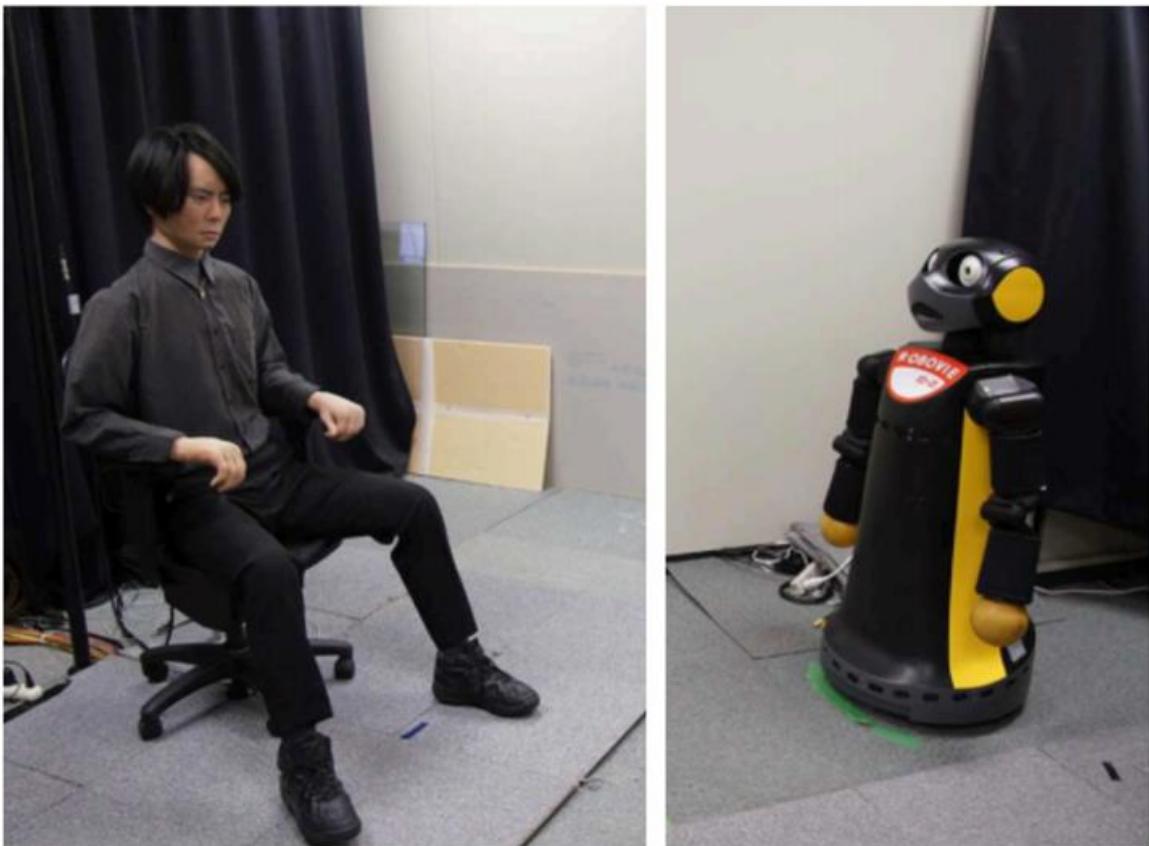


Рис. 3. Роботы Geminoid-2 (слева) Robovie-R2 (справа), использованные в исследовании Я. Злотовски и коллег. Иллюстрация из [Złotowski et al., 2015]

проекция соответствующего экспериментальному условию “лица” [Paetzl & Castellano, 2019]. В дальнейшем робот взаимодействовал с испытуемыми, рассказывая о себе и играя в игру “20 вопросов”, при этом демонстрируя мимику. Мы полагаем, что использование данной технологии ретропроекции позволит применить все преимущества 3D-морфов и возможности их анимации к изучению мультимодальных проявлений феномена 3Д. Отдельно следует отметить, что воздействование реальных роботов в исследовании эффекта зловещей долины может быть также перспективным с точки зрения изучения движущихся агентов, однако в данной работе, как и было сказано, мы не будем рассматривать эту исследовательскую область.

Размер континуума человекоподобия и количество экспериментальных условий

Последним вопросом, который необходимо рассмотреть в рамках обсуждения операционализации антропоморфности, является размер используемого в исследовании континуума, от которого зависит также количество экспериментальных условий, предъявляемых испытуемым. В обзоре С. Лэй и коллеги указывали на то, что минимальным количеством стимулов в исследовании является 5 штук, таким образом, что крайние положения занимают прототипные роботы и человек [Lay et al., 2016]. Мы согласны с тем, что это требование адекватно для воспроизведения “провала зловещей долины”, то есть негативной части исследуемого феномена. Однако, как и говорилось ранее, эффект ЗД подразумевает также и увеличение позитивного эмоционального ответа по мере роста антропоморфности агента перед наблюдаемым “провалом”, в связи с чем мы полагаем, что для более или менее точного воспроизведения графика М. Мори следует использовать континуум размером в 6-7 стимулов – таким образом, чтобы пик эмоционального ответа также мог быть отражён на графике. При этом, разумеется, использование как можно более продолжительного континуума человекоподобия значительно расширяет возможности моделирования кривой зловещей долины — это было продемонстрировано в исследовании М. Матур и соавторов, приблизившихся к воссозданию оригинальной кривой, хоть и не в ожидаемом виде [Mathur et al., 2020].

В соответствии с количеством используемых стимулов определяется и количество экспериментальных условий. Как правило, в исследованиях феномена ЗД используется внутригрупповой план, где каждому из испытуемых предъявляются все стимулы. При этом либо каждый из стимулов образует отдельное экспериментальное условие, либо одно условие включает несколько стимулов (например: Strait et al., 2017). Однако, как было указано в репликационном исследовании Дж. Паломаки и коллег, при небольшом

количестве стимулов внутригрупповой план может порождать систематические смешения, связанные с тем, что испытуемые имеют возможность сравнения эмоций, вызываемых у них различными агентами на одном континууме [Palomäki et al., 2018]. В связи с этим вслед за авторами можно предположить, что оптимальным решением будет либо использование в исследованиях ЗД межгруппового плана, устраняющего эти смешения, либо сохранение внутригруппового плана при значительном увеличении количества стимулов, делающим практически невозможным их субъективное сравнение.

Глава 3. Операционализация эмоционального ответа

В этой главе будут рассмотрены способы операционализации второго аспекта, определяющего феномен зловещей долины, а именно эмоционального ответа на предъявляемых агентов. Мы обсудим наиболее распространённые методы его измерения, такие как субъективная оценка и показатели когнитивного конфликта, и проанализируем другие методы, используемые сравнительно реже.

Измерение с помощью субъективной оценки

Самым первым способом операционализации эмоционального ответа были измерения, построенные на субъективной оценке. В связи с тем, что, как было отмечено в главе 1, предполагаемый М. Мори эмоциональный ответ является качественно специфическим и не имеет подходящих для его описания слов в английском (или другом) языке, за время исследований феномена было предложено множество вариантов определения того, какие именно эмоции должны вызывать агенты на континууме человекоподобия.

Дж. Катсири и коллеги, указывают на то, что одним из первых предложенных вариантов перевода и измерения был конструкт “знакомости” (familiarity), который позже был пересмотрен в пользу “привлекательности” (likeability) [Kätsyri et al., 2015]. Этот конструкт был включён в созданный для оценки процессов HRI опросник Годспид, разработанный К. Бартнеком и коллегами. Данный инструмент использует шкалы семантического дифференциала и помимо привлекательности позволяет измерять такие показатели, как антропоморфизм, одушевлённость, воспринимаемый интеллект и воспринимаемую безопасность агента, с которым взаимодействует человек [Bartneck et al., 2008]. Позже данный инструмент подвергся критике: М. Дестеф и коллеги указывали на то, что многие из используемых в опроснике шкал коррелируют друг с другом [Desteph et al., 2015], а Ч. Хо и К. МакДорман отметили, что валидность и надёжность инструмента не были эмпирически

проверены. Авторы также указали на отсутствие в опроснике Годспид такого показателя, как “жуткость” или “зловещесть”, который играет важную роль в определении негативного полюса эмоционального ответа. В связи с этим Ч. Хо и К. МакДорман был разработан альтернативный инструмент, также использующий семантический дифференциал и включающий показатели жуткости (eeriness), привлекательности (attractiveness), теплоты (warmth), и человечности (humanness) [Ho & MacDorman, 2010]. Однако и данный инструмент подвергся критике Дж. Катсири и коллег, указавших на нерелевантность некоторых пунктов опросника по отношению к измеряемому конструкту, на отсутствие в инструменте шкалы “знакомости”, которая может быть достаточно важной для понимания феномена зловещей долины и на то, что не все шкалы предложенного семантического дифференциала вообще можно применять к оценке роботов. Можно заметить, что на данный момент единого и признаваемого всеми инструмента для субъективной оценки вызываемых агентами эмоций не существует. Вместо этого многие исследователи продолжают предлагать собственные показатели для измерения эмоционального ответа, такие как “близость” (affinity) [Kätsyri et al., 2015], “приятность” (pleasantness), “отвращение” (disgust) [Palomäki et al., 2018] и так далее. Таким образом, точное определение специфики эмоций, вызываемых агентами в исследованиях эффекта ЗД, остаётся важным исследовательским вопросом.

Сейчас необходимо отметить, что при разработке конструктов, адекватно описывающих эмоциональный ответ на кривой ЗД нужно учитывать не только положительные измерения, то есть близость, привлекательность и другие, которые должны постепенно возрастать, снижаясь в “провале”, но и отрицательные измерения, то есть жуткость и отвращение. Согласно предсказаниям М. Мори их показатели должны оставаться нейтральными и повышаться одновременно с понижением положительных эмоций. Динамика обоих типов показателей является важной при исследовании феномена.

Измерение когнитивного конфликта

Альтернативы показателям, построенным на субъективной оценке, возникли в парадигме гипотезы категоризации. Они используются преимущественно исследователями в этой области с целью измерения когнитивного конфликта, возникающего при опознании неоднозначного агента в задании с вынужденным двухальтернативным выбором. В таком задании испытуемым предлагается отнести представленного агента к одной из двух категорий, и прежде чем мы перейдём к обсуждению конкретных показателей, подлежащих измерению, необходимо рассмотреть, какие именно категории обычно используются.

Несложно догадаться, что ясность в этом вопросе также отсутствует. Некоторые авторы предлагают использовать категории “человек” и “робот” (или другую категорию, соответствующую агенту на отрицательном полюсе человекоподобия, например, “животное”). Такая процедура использовалась в исследованиях Ю. Я마다 и соавторов [Yamada et al., 2013], М. Матур и соавторов [Mathur et al., 2020]. Другие авторы, например Е. Вайз и П. П. Вайс, предпочитают использовать ярлыки “человек” и “не человек”, таким образом не определяя категорию на отрицательном полюсе человекоподобия [Weis & Wiese, 2017; Wiese & Weis, 2020]. Сложно однозначно утверждать, какое из этих решений является предпочтительным, так как оно зависит от конкретных представлений авторов о природе феномена. Вероятно, с точки зрения эволюционной перспективы, где существо, не являющееся однозначно человеком, может быть рассмотрено как потенциально опасное и, следовательно, вызывать страх [Burleigh et al., 2013], не имеет значения, о каком существе идёт речь, а потому определение категории, противопоставляемой “человеку”, так же не слишком важно. Однако если в качестве источника негативных эмоций рассматривается когнитивный конфликт при категоризации как таковой, обе категории должны быть строго определены, в том числе во избежание смешения различных континуумов человекоподобия.

Теперь перейдём к рассмотрению тех показателей, которые используются в задании на категоризацию для измерения когнитивного конфликта. Первым из них является время реакции (ВР), увеличение которого свидетельствует о сложности опознания агента. Одновременно с ним современные исследования предлагают использовать технологию маустрекинга, так как показатели отклонения траектории мыши в задаче на категоризацию также могут служить валидной мерой когнитивного конфликта [Freeman & Ambady, 2010]. Например, в исследовании М. Матур и коллег использовались такие показатели, как: количество разворотов мыши; максимальное отклонение от идеальной траектории, соответствующей “однозначному” отнесению объекта к категории; площадь под траекторией отклонения; скорость курсора и общее время реакции [Mathur et al., 2020].

Рассмотрение того, насколько каждый из представленных показателей действительно может служить способом измерения когнитивного конфликта, будет опущено, поскольку данное обсуждение выходит за рамки исследуемой темы. Тем не менее, необходимо обратить внимание на то, является ли вообще адекватным связывание когнитивного конфликта с негативными эмоциями, наличие которых предполагает эффект зловещей долины. Дж. Катсири и коллеги указывали на то, что для полноценной проверки гипотезы когнитивного конфликта требуется выполнение ряда условий, в которые, помимо прочих, входит, во-первых, наличие сопровождающих когнитивный конфликт снижения положительных эмоций и повышения отрицательных, и, во-вторых, доказательство того, что эти эмоциональные изменения действительно вызываются сложностью различения стимулов на границе категории [Kätsyri et al., 2015]. К сожалению, как правило постулат о связи когнитивного конфликта с негативной эмоциональной реакцией принимается в качестве аксиомы, и описанные требования игнорируются исследователями. Тем не менее в будущих экспериментах они должны выполняться, например, путём одновременного фиксирования показателей когнитивного конфликта и эмоциональной оценки. В

таком случае следует проконтролировать смещения, связанные с последовательностью выполнения заданий, как это было сделано в исследовании К. МакДорман и Д. Чаттопадьи [MacDorman & Chattopadhyay, 2016].

Ещё одним недостатком гипотезы когнитивного конфликта в целом является отсутствие определения специфики вызываемых негативных эмоций. Существуют предположения о том, что конфликт сам по себе является их источником [Weis & Wiese, 2017] или что данные эмоции сопровождают процесс подавления когнитивного конфликта [Ferrey et al., 2015], однако ни одно из них не объясняет, почему люди переживают специфические ощущения отвращения или страха. Более того, гипотеза категоризации не объясняет, почему по мере увеличения человекоподобия до и после провала увеличиваются положительные эмоции. Связано ли восприятие однозначных и знакомых категорий с переживанием “близости” или “теплоты”? Если да, то каким образом? На эти вопросы необходимо ответить и проверить соответствующие предположения экспериментально, чтобы гипотеза когнитивного конфликта могла претендовать на объяснение феномена зловещей долины.

Альтернативные способы измерения

В заключительном разделе главы будут рассмотрены некоторые менее популярные способы измерения эмоционального ответа и других зависимых переменных, используемых в исследованиях феномена зловещей долины. Одним из таких способов можно считать предложенное С. Вонгом и П. Рошатом задание “визуального приближения” (visual looming task). В данном задании испытуемым демонстрируется фотография лица агента, резко приближающегося к ним до достижения определённого размера и затем исчезающего. Участников эксперимента просят нажать на кнопку, когда, по их мнению, данное лицо соприкоснулось бы с их лицом, если бы продолжало движение. Авторы утверждают, что систематическая переоценка времени, необходимого для “столкновения” лиц, может свидетельствовать о восприятии агентов как

неприятных и устрашающих, вызывая желание преждевременно предотвратить контакт с ними [Wang & Rochat, 2017]. Применение этого задания в эксперименте не было успешным - вероятно, из-за неправильного формулирования инструкции, подтолкнувшей испытуемых к рассуждению о скорости лиц и вероятности контакта, нежели быстрой оценке. Однако использование задания визуального приближения в исследованиях ЗД может оказаться перспективным, так как оно может служить объективной альтернативой оценкам “жуткости”, построенным на самоотчёте.

В главе 1 было отмечено, что одной из гипотез возникновения эффекта является страх смерти, вызываемый похожими на мёртвые тела андроидами. Несмотря на то, что эта точка зрения не получила большого распространения, в некоторых исследованиях данное предположение проверялось путём измерения доступности в памяти ассоциаций о смерти [Koschate et al., 2016]. Испытуемым после взаимодействия со “зловещими” агентами предлагалось задание на дополнение слов с неоднозначными окончаниями (“гриб/гроб”, “смерть/смех”). Таким образом, гипотеза страха смерти имеет способ измерения, хотя он не лишён недостатков. Например, необходимо измерять изначальную доступность ассоциаций о смерти, чтобы отслеживать её изменение после экспериментального воздействия. Кроме того, при таком способе измерения установить каузальную связь между доступностью ассоциаций и негативной оценкой агентов по-прежнему невозможно, что осложняет проверку гипотезы страха смерти.

Ещё одним нестандартным способом измерения отношения к роботам являются тесты ассоциаций, в частности методика Краткий тест имплицитных ассоциаций (Brief Implicit Association Test, BIAT) [Sriram & Greenwald, 2009], использованная в исследовании Я. Злотовски и коллег для измерения воспринимаемой “жуткости” роботов [Złotowski et al., 2015]. Такие тесты, несомненно, могут служить альтернативой методам эксплицитной субъективной оценки, хотя в большей степени их использование будет полезным при

измерении у испытуемых негативных стереотипов по отношению к роботам, о чём будет сказано в следующей главе.

Более успешным применение рассмотренных выше методик может оказаться в парадигме социального функционализма. Так, А. Оливера ла-Роса, обсуждая гипотезу избегания психопатов в качестве потенциального объяснения природы эффекта зловещей долины, указывает на возможность применения тестов имплицитных ассоциаций для проверки этого предположения [Olivera-La Rosa, 2018]. Можно также предположить, что в рамках данной теории уместно использование других методов исследований, применяемых обычно в социальной психологии. Отталкиваясь от предположения о том, что роботы воспринимаются как агенты с неопределенным и потенциально негативным моральным статусом, доверие к ним и степень воспринимаемой предсказуемости их поведения можно проверить путём моделирования ситуаций социальных дilemm. Такая попытка уже предпринималась в более раннем исследовании М. Матур и Д. Ричлинга, где испытуемые принимали участие в “Игре инвестиций” из теории игр. Испытуемым предлагалось определить, какое количество денег они готовы отдать роботу, ожидая, что определённая сумма будет им возвращена [Mathur & Reichling, 2016]. В эксперименте использовался только первый раунд игры, отданная роботу сумма определяла уровень доверия по отношению к нему. Предположительно, данное исследование можно было бы расширить сверх одного раунда игры с целью изучения динамики отношения к роботу по мере получения информации о его поведении. Также можно было бы использовать и другие социальные дилеммы, например, классическую дилемму узника или дилемму “труса” (“chicken dilemma”), в которой выигрыш обоих участников возможен только в случае, если один из них отступится от своей изначальной позиции [Van Lange et al., 2013]. Таким образом, поведение испытуемых в этой дилемме может отражать их представления о просоциальной направленности робота и, соответственно, их моральном статусе.

Также говоря о применении социально-психологического инструментария в исследованиях феномена ЗД можно предложить использование выделенных С. Фиске и коллегами универсальных измерений социального познания, то есть оценок представителей аутгруппы по шкалам теплоты и компетентности [Fiske et al., 2007]. Данные характеристики совместно с некоторыми параметрами опросника Годспид легли в основу специализированного инструмента для оценки HRI - Опросника социальных характеристик роботов (Robotic Social Attributes Scale, RoSAS), разработанного К. Карпинелла и соавторами [Carpinella et al., 2017]. Данный инструмент уже применялся в исследовании М. Пейтсел и Дж. Кастеллano в контексте фиксирования динамики изменений в восприятии робота по мере взаимодействия человека с ним [Paetzel & Castellano, 2019]. Тем не менее, этот опросник или некоторые его шкалы вполне могут служить хорошей альтернативой измерению классических показателей “близости” или “привлекательности” при рассмотрении эффекта ЗД с точки зрения социального функционализма.

Глава 4. Направления развития исследований

В предыдущих главах были освещены основные методологические проблемы исследований в области феномена зловещей долины и способы их решения с точки зрения различных подходов. Сейчас мы считаем нужным обсудить перспективы развития всей сферы, указав на возможные направления работы, такие как изучение индивидуальных различий, применение аппаратных методов и усовершенствованное математическое моделирование эффекта.

Изучение индивидуальных и возрастных различий

Исследование индивидуальных различий в чувствительности к эффекту зловещей долины представляется нам перспективным с точки зрения как развития понимания самого эффекта, так и совершенствования дизайна исследований.

На данный момент существует несколько работ в этой области. Например, в исследовании К. МакДорман и С. Энтеzари было показано, что такие черты, как нейротизм и тревожность усиливают вероятность возникновения эффекта ЗД, а негативное отношение к роботам, также способствующее его проявлению, может быть вызвано религиозным фундаментализмом и опосредоваться наличием негативных аттитюдов [MacDorman & Entezari 2015]. В другом исследовании, проведённом М. Оканда и коллегами, было установлено, что склонность к анимистическим верованиям увеличивает вероятность атрибуции роботам социальных характеристик, так как способствует восприятию их как одушевлённых агентов [Okanda et al., 2019]. Данные результаты могут косвенно свидетельствовать в пользу теории социального функционализма в качестве объяснения возникновения феномена, поскольку выделенные авторами черты в большей степени являются приобретёнными и связаны с социальной оценкой агентов. В то же время исследование конституциональных особенностей нервной системы, усиливающих эффект ЗД, может быть важным с точки зрения

когнитивных теорий объяснения феномена. Другим направлением исследований в области эффекта ЗД может стать изучение различий в чувствительности к нему у людей, имеющих разный опыт взаимодействия с потенциально “зловещими” агентами. Ввиду того, что многие люди имеют опыт игры в компьютерные игры и взаимодействия с виртуальными персонажами, это может сказываться на их подверженности эффекту.

Развитие этого направления может быть полезным из соображений дизайна экспериментов. Так, М. Читам и соавторы в своём исследовании во избежание смешений контролировали то, чтобы у испытуемых не было большого опыта игры в компьютерные игры или 3D-моделирования персонажей [Cheetham et al., 2013], хотя достоверной информации о том, какое влияние эти факторы оказывают на экспериментальные результаты, ещё не было получено. Контроль смешений, вызванных личностными чертами, в свою очередь, можно произвести с помощью вышеупомянутых методик измерения имплицитных ассоциаций (например, ВIАТ) или других психометрических инструментов.

Важным, по нашему мнению, также может быть исследование возрастных различий в чувствительности к эффекту зловещей долины. На данный момент результаты в этой области противоречивы. Ранние работы указывают на то, что дети становятся подверженны эффекту в очень раннем возрасте 12 месяцев [Lewkowicz & Ghazanfar, 2012], однако данные результаты опровергаются в более современных исследованиях. Так, К. Бринк и коллеги установили, что дети становятся чувствительными к эффекту ЗД в возрасте 9 лет и старше, одновременно с приобретением способности к атрибуции агенту сознания [Brink et al., 2019]. Существуют также свидетельства того, что дети с расстройствами аутистического спектра не подвержены переживанию эффекта зловещей долины [Feng et al., 2018]. На основании этих данных можно предположить, что исследование феномена ЗД в перспективе его развития в онтогенезе будет ценным для понимания причин и механизмов его возникновения.

Применение аппаратных методов

Парадигма когнитивного конфликта позволила исследователям феномена зловещей долины перейти от рассмотрения субъективной эмоциональной оценки к регистрации более объективных показателей сложности категоризации. Тем не менее, множество объективно регистрируемых индикаторов не исчерпывается ими, в связи с чем мы рассмотрим другие аппаратные методы, которые могут быть применены исследователями.

Достаточно перспективным, может быть применение технологии ай-трекинга. Результаты исследования М. Читам и коллег свидетельствуют о том, что при выполнении задания на категоризацию неоднозначных агентов наиболее важными областями, на которых испытуемые фиксируются больше всего, являются глаза и нос [Cheetham et al., 2013]. Эти данные могут быть полезны для сторонников гипотезы несоответствия перцептивных признаков, поскольку указывают на наиболее важные для категоризации лица признаки, конфликт которых может порождать негативный эмоциональный ответ. Также с помощью технологии ай-трекинга можно было бы проверить наличие качественных отличий в паттернах изучения лиц, занимающих различные положения на континууме человекоподобия, и установить связь этих паттернов с эмоциональным ответом, что способствовало бы лучшему пониманию феномена.

Другим направлением использования аппаратных методов является их применение с целью регистрации эмоционального ответа. По нашим сведениям, не существует работ в области эффекта зловещей долины, в которых для фиксирования динамики эмоционального ответа использовались бы показатели электроэнцефалограммы (ЭЭГ) или кожно-галванической реакции (КГР). Однако современные исследования показывают, что опираясь на показатели КГР можно дифференцировать такие негативные эмоции, как страх и отвращение [Wu et al., 2010]. Такой подход может пролить свет на специфическое содержание эмоций, испытываемых при контакте со “зловещими” агентами, а

также помочь отследить изменения в эмоциональном состоянии испытуемого. Для исследования динамики переживаний существуют и более современные способы оценки, включающие в себя помимо КГР измерение показателей сердцебиения и ЭЭГ [Val-Calvo et al., 2020]. Данные технологии могут послужить альтернативой или дополнением к субъективной оценке испытуемых, обычно служащей единственным индикатором эмоционального ответа.

Говоря об аппаратных методах, необходимо обсудить применение методов нейровизуализации в исследованиях феномена ЗД. Они использовались в работе М. Читам и Л. Джанке, в которой было установлено, что зонами мозга, реагирующими активацией на изменения и, в частности, категориальный переход на континууме человекоподобия, являются амигдала, веретеновидная извилина, островок и гиппокамп [Cheetham & Jancke, 2013]. Эти данные не являются достаточными для однозначного установления нейронных коррелятов эффекта зловещей долины, однако показывают, что такая процедура вполне возможна. С учётом того, что как теория рассогласованности перцептивных признаков, так и теория когнитивного конфликта предполагает наличие нейронных коррелятов соответствующих процессов (например, сторонники первой гипотезы утверждают, что несогласующиеся признаки порождают соперничество кодирующих их нейронных популяций [MacDorman & Chattopadhyay, 2016]), изучение паттернов мозговой активации, возникающей при восприятии агентов на континууме человекоподобия, может быть перспективным для обоих направлений.

Математическое моделирование эффекта

В главе 2 при обсуждении размера используемого в исследованиях континуума антропоморфности был затронут вопрос моделирования оригинальной кривой “зловещей долины”. Сейчас мы считаем нужным раскрыть эту тему, поскольку единое мнение о возможности воспроизведения оригинального графика М. Мори отсутствует.

В обзоре С. Лэй и коллег было указывано, что воссоздание графика ЗД подразумевает наличие линейной зависимости между параметрами человекоподобия и эмоционального ответа с единственным отрицательным отклонением от данной зависимости в диапазоне от 50% до 100% антропоморфности [Lay et al., 2016]. Хотя данное описание является верным, оно не представляет собой никаких практических рекомендаций к моделированию кривой, являясь скорее параметром оценки полученного эмпирического графика.

Комплексная математическая модель, позволяющая не только описать график, но и предсказать его форму, была предложена Р. К. Муром и получила название байесовской модели зловещей долины [Moore, 2012]. В данной модели переменными, предсказывающими валентность эмоционального ответа являются вероятность возникновения знакомого стимула и вероятность возникновения эмоционального напряжения. Первый параметр определяется распределением среди стимулов перцептивных признаков или “подсказок”, соответствующих “знакомой” (человек) и “малознакомой” (робот/животное) категориям. Второй параметр зависит от того, как соотносятся перцептивные признаки двух категорий - в случае, если переход от одной категории к другой является резким, и перцептивные признаки конфликтуют, эмоциональное напряжение на границе категорий будет наибольшим, а в случае плавного перехода между категориями напряжение может не наблюдаться вообще, не оказывая воздействия на валентность эмоционального ответа. Таким образом, с помощью данной модели, отталкиваясь только от параметров распределения перцептивных признаков, можно не только предсказать возникновение негативного эмоционального ответа, но и его силу, а также общую форму кривой.

Несмотря на то, что некоторые авторы утверждают, что полученные ими результаты соотносятся с байесовской моделью (например: Broadbent et al., 2013), нам неизвестны исследования, которые на этапе подбора стимульных материалов контролировали бы параметры распределения признаков

исключительно с целью моделирования кривой, в связи с чем этот вопрос будет необходимо решить в будущих исследованиях. Такой подход, несомненно, имеет свои недостатки - например, он не объясняет возникновение качественно своеобразных негативных эмоций. Тем не менее, его применение может не только приблизить исследователей к пониманию эффекта ЗД, но и разрешить противоречия между гипотезами перцептивного несоответствия и категоризации, поскольку в байесовской модели категориальное восприятие зависит непосредственно от имеющихся перцептивных признаков.

Заключение

Резюмируем результаты проделанной работы. Обсудив главные методологические проблемы, с которыми сталкиваются исследователи в области феномена зловещей долины, мы рассмотрели возможные способы их решения, а также, проведя их критический анализ, предложили усовершенствования. Было установлено, что наиболее успешной и универсальной техникой подбора стимульных материалов может быть создание морфов фотoreалистичных 3D-моделей. Однако нам не удалось выделить наиболее адекватный способ измерения эмоционального ответа и других зависимых переменных в экспериментальных исследованиях ЗД, в связи с чем были предложены способы улучшения имеющихся - например, применение социально-психологического инструментария для оценки роботов как социальных агентов. Также было указано на основные направления дальнейшей работы в области, к которым относится как совершенствование методических процедур, - применение аппаратных методов и более прогрессивных способов математического моделирования - так и теоретическое развитие исследований, способствовать которому может изучение феномена ЗД в перспективе психологии развития. Мы полагаем, что проведённый критический анализ методологии исследований феномена зловещей долины и предложенные рекомендации помогут исследователям приблизиться к лучшему пониманию изучаемого эффекта.

Список источников

- Bartneck, C., Kulić, D., Croft, E., Zoghbi, S., (2008). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, v. 1(1), pp. 71-81
- Birks, M., Bodak, M., Barlas, J., Harwood, J., Pether, M., (2016). Robotic seals as therapeutic tools in an aged care facility: a qualitative study. *Journal of Aging Research*, 1–7. doi: 10.1155/2016/8569602
- de Borst, A. W., de Gelder, B., (2015). Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective, *Frontiers in Psychology*, 5:576
- Brink K.A., Gray K., Wellman H.M., (2019). Creepiness Creeps In: Uncanny Valley Feelings Are Acquired in Childhood, *Child Development*, v. 90(4), pp. 1202-1214
- Broadbent, E., Kumar, V., Li, X., Sollers, J. 3rd, Stafford, R.Q., MacDonald, B.A., Wegner, D.M, (2013). Robots with Display Screens: A Robot with a More Humanlike Face Display Is Perceived To Have More Mind and a Better Personality
- Burleigh, T.J., Schoenherr, J.R., Lacroix, G.L., (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces, *Computers in Human Behaviour*, v. 29 (3), pp. 759–771
- Burleigh, T.J., Schoenherr, J.R., (2015). A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization? *Frontiers in Psychology*, 5:1488

Carpinella, C.M., Wyman, A.B., Perez, M.A., Stroessner, S.J. (2017). The Robotic Social Attributes Scale (RoSAS): Development and Validation, ACM/IEEE International Conference on Human-Robot Interaction, Part F127194, pp. 254-262

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., Jancke, L., (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data, *Frontiers in psychology*, 4:108

Cheetham, M., Jancke, L. (2013). Perceptual and Category Processing of the Uncanny Valley Hypothesis' Dimension of Human Likeness: Some Methodological Issues, *Journal of Visualized Experiments*. (76), e4375, doi:10.3791/4375

Church, W., Ford, T., Perova, N., Rogers, C. (2010). Physics with robotics: using Lego Mindstorms in high school education, *AAAI Spring Symposium - Technical Report SS-10-03*, pp. 47-49

Destephe, M., Brandao, M., Kishi, T., Zecca, M., Hashimoto, K., Takanishi, A., (2015). Walking in the uncanny valley: importance of the attractiveness on the acceptance of a robot as a working partner, *Frontiers in psychology*, 6:204

Felnhofer, A., Kaufmann, M., Atteneder, K., Kafka, J.X., Hlavacs, H., Beutl, L., Hennig-Fast, K., Kothgassner, O.D., (2018). The mere presence of an attentive and emotionally responsive virtual character influences focus of attention and perceived stress, *International Journal of Human Computer Studies*, v. 132, pp. 45-51

Feng, S., Wang, X., Wang, Q., Fang, J., Wu, Y., Yi, L., Wei, K., (2018). The uncanny valley effect in typically developing children and its absence in children with autism spectrum disorders, *PLoS ONE*, 13:e0206343

Ferrey, A.E., Burleigh, T.J., Fenske, M.J., (2015). Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley, *Frontiers in Psychology*, 6:249

Fiske, S.T., Cuddy, A.J.C., Glick, P. (2007). Universal dimensions of social cognition: warmth and competence, *Trends in Cognitive Sciences*, v. 11 (2), pp. 77-83

Freeman, J.B., Ambady, N., (2010). MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method, *Behavioral Response Methods*, v. 42 (1), pp. 226–241.

Ishak, N.I., Yusof, H.M., Sidek, S.N., Rusli, N., (2019). Robot selection in robotic intervention for ASD children, 2018 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2018 - Proceedings 8626679, pp. 156-160

Haslam, N., (2006). Dehumanization: an integrative review, *Personality and Social Psychology Review*, v. 10, pp. 252–264. doi: 10.1207/s15327957pspr1003/4

Ho, C.-C., MacDorman, K.F., (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices, *Computers in Human Behavior*, v. 26(6), pp. 1508-1518

Jentsch, E. (1997). On the psychology of the uncanny (1906). *Angelaki: Journal of the Theoretical Humanities*, 2(1), 7–16.

Kätsyri, J., Förger, K., Mäkäräinen, M., Takala, T., (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, 6:390

Kawabe, T., Sasaki, K., Ihaya, K., Yamada, Y. (2017). When categorization-based stranger avoidance explains the uncanny valley: A comment on MacDorman and Chattopadhyay (2016). *Cognition*, v. 161, pp. 129–131

Koschate, M., Potter, R., Bremner, P., Levine, M., (2016). Overcoming the uncanny valley: Displays of emotions reduce the uncanniness of humanlike robots, ACM/IEEE International Conference on Human-Robot Interaction 2016-April, 7451773, pp. 359-365

Lay, S., Brace, N., Pike, G., Pollick, F., (2016). Circling around the uncanny valley: Design principles for research into the relation between human likeness and eeriness, i-Perception 7(6), 2041669516681309

Lewkowicz D. J., Ghazanfar A. A. (2012). The development of the uncanny valley in infants. Developmental Psychobiology, 54(2), 124–132

Mäkäräinen, M., Kätsyri, J., Takala, T., (2014). Exaggerating facial expressions: a way to intensify emotion or a way to the uncanny valley? Cognitive Computation, v.6, pp. 708–721. doi: 10.1007/s12559-014-9273-0

MacDorman, K., (2005). Androids as an experimental apparatus: why is there an uncanny valley and can we exploit it?, Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop, Stresa, 106–118.

MacDorman, K.F., (2005). Mortality salience and the uncanny valley, 5th IEEE-RAS. IEEE, pp. 399–405.

MacDorman, K.F., Green, R.D., Ho, C.-C., Koch, C. (2009). Too real for comfort: Uncanny responses to computer generated faces. Computers in Human Behavior, 25(3), 695–710. <http://dx.doi.org/10.1016/j.chb.2008.12.026>.

MacDorman, K.F., Entezari, S.O., (2015) Individual differences predict sensitivity to the uncanny valley, Interaction Studies, v. 16(2), pp. 141-172

MacDorman, K.F., Chattopadhyay, D., (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not, Cognition, v. 146, pp. 190–205

Mathur, M. B., Reichling, D.B., (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley, *Cognition*, v. 146, pp. 22-32

Mathur, M. B., Reichling, D.B., Lunardini, F., Geminiani, A., Antonietti, A., Ruijten, P.A.M., Levitan, C.A., Nave, G., Manfredi, D., Bessette-Symons, B., Szuts, A., Aczel, B., (2020). Uncanny but not confusing: Multisite study of perceptual category confusion in the Uncanny Valley, *Computers in Human Behavior*, v. 103, pp. 21–30

Mitchell, W.J., Szerszen, K.A., Lu, A.S., Schermerhorn, P.W., Scheutz, M., MacDorman, K.F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception* 2, 10–12. doi: 10.1068/i0415

Mori, M., (1970). The uncanny valley, *Energy*, v. 7(4), pp. 33–35

Mori, M., MacDorman, K.F., Kageki, N., (2012). The Uncanny Valley, *IEEE Robotics & Automation Magazine*, v. 19(2), pp. 98-100, doi: 10.1109/MRA.2012.2192811.

Moore, R.K., (2012). A Bayesian explanation of the ‘Uncanny Valley’ effect and related psychological phenomena, *Scientific reports*, 2:864

Mousas, C., Anastasiou, D., Spantidi, O., (2018). The effects of appearance and motion of virtual characters on emotional reactivity, *Computers in Human Behavior*, v. 86, pp. 99-108

Okanda, M., Taniguchi, K., Itakura, S., (2019) The role of animism tendencies and empathy in adult evaluations of robot, *HAI 2019 - Proceedings of the 7th International Conference on Human-Agent Interaction*, 25 September 2019, pp. 51-58

Olivera-La Rosa, A., (2018). Wrong outside, wrong inside: A social functionalist approach to the uncanny, *New Ideas in Psychology*, v. 50, pp. 38-47

Paetzl, M., Castellano, G., (2019). Let me get to know you better: Can interactions help to overcome uncanny feelings? HAI 2019 - Proceedings of the 7th International Conference on Human-Agent Interaction, pp. 59-67

Palomäki, J., Kunnari, A., Drosinou, M., Koverola, M., Lehtonen, N., Halonen, J., Repo, M., Laakasuo, M., (2018). Evaluating the replicability of the uncanny valley effect, *Heliyon*, 4:e00939

Reategui, E., Boff, E., Campbell, J.A., (2008). Personalization in an interactive learning environment through a virtual character, *Computers & Education*, v. 51(2), pp. 530-544

Sriram, N., Greenwald, A. G., (2009). The brief implicit association test, *Experimental Psychology*, v. 56(4), pp. 283–294. doi: 10.1027/1618-3169.56.4.283

Strait, M.K., Floerke, V.A., Ju, W., Maddox, K., Remedios, J.D., Jung, M.F., Urry, H.L., (2017). Understanding the Uncanny: Both Atypical Features and Category Ambiguity Provoke Aversion toward Humanlike Robots, *Frontiers in Psychology*, 8:1366

Takayama, L., Ju, W., Nass, C. (2008). “Beyond dirty, dangerous and dull: what everyday people think robots should do,” in Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, Amsterdam, pp. 25–32.

Triebel, R., Arras, K., Alami, R., Beyer, L., Breuers, S., Chatila, R., et al. (2016). “Spencer: a socially aware service robot for passenger guidance and help in busy airports,” *Springer Tracts in Advanced Robotics* 113, pp. 607-622

Val-Calvo, M., Álvarez-Sánchez, J.R., Ferrández-Vicente, J.M., Díaz-Morcillo, A., Fernández-Jover, E., (2020). Real-Time Multi-Modal Estimation of Dynamically Evoked Emotions Using EEG, Heart Rate and Galvanic Skin Response, *International Journal of Neural Systems*, v. 30(4), 2050015

Van Lange, P. A.M, Joireman, J., Parks, C.D., Van Dijk, E., (2013). The psychology of social dilemmas: A review, *Organizational Behavior and Human Decision Processes*, v. 120 (2), pp. 125-141

Volante, M., Babu, S.V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., Daily, S. B., Fasolino, T., (2016). Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations, *IEEE Transactions on Visualization and Computer Graphics* 22(4), 7383334, pp. 1326-1335

Wang, S., Rochat, P., (2017). Human Perception of Animacy in Light of the Uncanny Valley Phenomenon, *Perception*, v. 46(12), pp. 1386–1411

Weis, P.P., Wiese, E., (2017). Cognitive conflict as possible origin of the uncanny valley, *Proceedings of the Human Factors and Ergonomics Society 2017-October*, pp. 1599-1603

Wiese, E., Metta, G., Wykowska, A., (2017). Robots As Intentional Agents: Using Neuroscientific Methods to Make Robots Appear More Social, *Frontiers in Psychology*, 8:1663

Wiese, E., Weis, P.P., (2020). It matters to me if you are human - Examining categorical perception in human and nonhuman agents, *International Journal of Human-Computer Studies*, v. 133, pp. 1–12

Wu, G., Liu, G., Hao, M. (2010). The Analysis of Emotion Recognition from GSR Based on PSO. In *Proceedings of the International Symposium on Intelligence Information Processing and Trusted Computing*, Huanggang, China, 28–29 October 2010; pp. 360–363.

Yamada, Y., Kawabe, T., Ihaya, K., (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon, *Japanese Psychological Research*, v. 55(1), pp. 20-32

Złotowski, J.A., Sumioka, H., Nishio, S., Glas, D.F., Bartneck, C., Ishiguro, H., (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception, *Frontiers in psychology*, 6:883