

УДК 159.9.072.59

В. А. Филимонов<sup>1</sup>

Институт математики им. С.Л. Соболева Сибирского отделения Российской Академии наук  
Новосибирск, Россия

В. С. Чернявская<sup>2</sup>

Владивостокский государственный университет экономики и сервиса  
Владивосток, Россия

### **Катастрофы субъективных оценок и их модели: к вопросу инструментализации диагностики эффекта «зловещей долины»**

Анализируются исследования феномена «зловещей долины», который проявляется в отсутствии монотонной зависимости симпатии к роботу от степени его антропоморфности. Приводятся примеры интерпретаций феномена с позиций разных подходов. Рассмотрены исследования данного эффекта как проявления и отношения к антропоморфности роботов, как проблемы идентичности, преодоления органичности, «протезности сознания». Описаны трудности и опыт измерения субъективных оценок: психосемантических методов, экспертных оценок. Представлены варианты анализа субъективных оценок, которые являются исключительно сложным феноменом. Проанализирован опыт использования разработанных методик исследования и диагностики феномена «зловещей долины», анализируются недостатки методик, показана ограниченность полученных результатов. Рассматриваются феномены бистабильности восприятия в контексте эффекта «зловещей долины». Констатируется наличие гистерезиса как причины таких феноменов. Явление гистерезиса подразумевает зависимость поведения системы от её предыстории. Обосновывается методика диагностики феномена «зловещей долины» с помощью фотороботов с разным уровнем человекоподобия и шкалами оценки и двумя вариантами диагностики в соответствии с обходами точки бифуркации, фотографий реальных роботов, игрушек и персонажей фильмов. Фотографии были отобраны таким образом, чтобы организаторы эксперимента могли классифицировать их по шкалам степени человекоподобия и угрозы, а также размещать по одной фотографии в каждой позиции матрицы эксперимента,

---

<sup>1</sup> Филимонов Вячеслав Аркадьевич – ст. науч. сотрудник, д-р техн. наук, профессор; e-mail: [trizkin@yandex.ru](mailto:trizkin@yandex.ru)

<sup>2</sup> Чернявская Валентина Станиславовна – д-р пед. наук, профессор, профессор кафедры философии и юридической психологии; e-mail: [valstan13@mail.ru](mailto:valstan13@mail.ru)

названной нами Матрица-25. Предложены два варианта проведения диагностической процедуры: левый обход (L): сначала снижается степень сходства с человеком, затем увеличивается угроза; правый обход (R): сначала увеличивается угроза, затем уменьшается степень сходства с человеком.

**Ключевые слова и словосочетания:** субъективные оценки, теория катастроф, гистерезис, рефлексия, организация экспериментов.

V.A. Filimonov

S.L. Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences  
Novosibirsk. Russia

V.S. Chernyavskaya

Vladivostok State University of Economics and Service  
Vladivostok. Russia

## **Disasters of subjective assessments and their models: to the question of instrumentalization of diagnostics of the "uncanny valley" effect**

The research of the "uncanny valley" phenomenon is analyzed. The phenomenon manifests itself in the absence of a monotonous dependence of sympathy for the robot on the degree of its anthropomorphism. Examples of interpretations of the phenomenon in the context of different approaches are given. A number of studies of this effect as a manifestation and attitude to the anthropomorphism of robots, as problems of identity, overcoming organicity, and "prosthetics of consciousness" are considered. The difficulties and experience of measuring subjective assessments are described: psychosemantic methods, expert assessments. The article presents the variants of the analysis of subjective assessments, which are an extremely complex phenomenon, analyzes the experience of using the developed methods of research and diagnosis of the "uncanny valley" phenomenon, analyzes the shortcomings of the methods (questionnaires), and shows the limitations of the results obtained. The phenomena of bistability of perception in the context of the "uncanny valley" effect are considered. The presence of hysteresis is stated as the cause of such phenomena. Recommendations on the organization of experiments in accordance with this hypothesis are given. The article substantiates the method of diagnosing the "uncanny valley" phenomenon using photographs of robots with different levels of human-like behavior and assessment scales and two diagnostic options, in accordance with the bifurcation point traversals. photos of real robots, toys, and movie characters. The photos were selected in such a way that the organizers of the experiment could classify them according to the scales of the degree of human likeness and threat, as well as place one photo in each position of the experiment matrix, which we called Matrix-25. Two variants of the diagnostic procedure are proposed: left bypass (L): first, the degree of similarity to the person decreases, then the threat increases; right bypass (R): first, the threat increases, then the degree of similarity to the person decreases.

**Keywords:** subjective assessments, catastrophe theory, hysteresis, reflection, organization of experiments.

### **Введение**

Прошло сорок три года с момента первого описания Масахири Мори феномена «зловещая долина» [10], но этого оказалось недостаточно для создания инструмента для исследования его в рамках психологии и смежных наук. Возможно, данный феномен отражает те грани человеческой реальности, которые не существовали ранее. Исследователи связывают данный эффект с недопониманием себя, с недостатком рефлексивных ресурсов [2]. Робототехника программирует, схематизирует и преодолевает органичность человека, в то время как человеку присуща способность быть мерой и критерием всех вещей. Антропоморфность искусственных предметов пугает человека, поскольку разрушает его целостность, единство, а возможно, имеет родовые источники, связанные с сохранением безопасности человека.

### **Предмет и источниковая база исследования, противоречия в имеющихся исследованиях и авторская позиция**

Субъективные оценки экспертного характера отличаются прогностичностью относительно отбора и профессиональной деятельности, однако точность этих оценок не всегда соответствует требованиям объективности.

О.А. Ананьина вслед за П. Уоттсом связывает эффект «зловещей долины» с метафорой протеза. Рассмотрение наблюдателем данного эффекта попадает в контекст самого эффекта. Автор пишет, что «протезная инаковость становится частью идентичности и угрожает субъектности пользователя» [1, с. 26]

Автор отмечает, что «выбраться из впадины «зловещей долины» можно лишь при реализации стратегии служения, бытия полезным, такие стратегии возможны, преимущественно, в профессиональной деятельности» [1]. Ценными для персонажей с «протезной идентичностью» являются ситуации, которые не просто затрагивают, но вызывают боль, гнев и другие сильные негативные аффекты, которые идентифицируют человека с собой (собирают) и витальностью.

В рамках психологии субъективной семантики В.П. Серкиным и другими учеными описаны структуры субъективного опыта личности [7]. Вариант исследования субъективного переживания отношения к изображениям роботов может объективизироваться в рамках анализа и систематизации ключевых аспектов индивидуальных субъективных данных. Важным здесь является то, что результаты соотнесения субъективного опыта с визуальными формами не зависят от языка, что позволяет объективизировать ключевые аспекты человеческого опыта в рамках разных языков и культур. Перспективы разработки такого варианта исследования требуют специальной разработки в рамках особой методологии.

Исследования эффекта «зловещей долины» проводились с помощью разного рода анкет. В частности, Ю.С. Орлова с соавторами решали задачу выбора наиболее приемлемого интерфейса социального робота, что привело авторов к выводу о том, что испытуемым, негативно относящимся к интерфейсу

социального робота андроидного типа, действительно присущи тревожность и эмпатия [4].

Л.Э. Семенова с соавторами использовали анкету с семью фото- и видеоматериалами, изображающими человекоподобных андроидов [6]. Стимульный материал сопровождался шкалами комфортности и списком для выбора эмоций. Выводы исследователей сводились к подтверждению гипотезы о снижении степени комфортности в восприятии роботов с возрастом. Оказалось, что спектры эмоций лиц разного возраста в восприятии роботов схожи, однако у взрослых испытуемых наиболее выражена эмоция интереса. В итоге исследователи указывают, что наличие данного эффекта препятствует внедрению роботов в сферу здравоохранения, а также отмечают недостаточность исследовательских данных об эффекте «зловещей долины».

На наш взгляд, методологические основы подхода к исследованию феномена «зловещей долины» обязательно должны включать системный подход, поскольку мы имеем дело со сложной системой. Особенно важно учитывать его, поскольку в настоящее время рассматриваются варианты моделирования поведения не только с роботами, но и с коллективом роботов [9].

Варианты анализа субъективных оценок являются исключительно сложным феноменом. Даже в таких сравнительно простых случаях, как оценка тяжести либо размеров предметов, можно зафиксировать конформизм, искажение при категоризации и т.п. Анализ способов организации экспериментов и учёт рефлексии позволяют дать удовлетворительное объяснение ряду феноменов. Ярким примером является обобщение В.А. Лефевром моделей Вебера-Фехнера и Стивенса [3].

На наш взгляд, количество типов роботов существенно превышает число стимулов, которое использовано в указанном исследовании, что должно быть учтено в разработке диагностической методики.

Мы рассматриваем эксперименты, в которых наблюдается явление гистерезиса, т.е. зависимость поведения системы от её предыстории. В технике одним из наиболее простых примеров служит триггер Шмитта. В психологии известным примером является бистабильность восприятия, когда заключение об интерпретации изображения зависит от последовательности предъявления серии изображений. Приведём простой авторский пример бистабильности интерпретации следующей последовательностью слов:

List, word, stone, gun, *cop*, сера, роса, хор, оса.

При чтении слева направо слово «*cop*» читается [*kop*] как англоязычное слово и интерпретируется как «*полицейский*», а при чтении справа налево как слово [*sor*] русского языка, что интерпретируется как «*мусор*».

Иногда в результатах экспериментов наличие гистерезиса не распознаётся, а феномены нарушения монотонности оценок констатируются, но не объясняются. Примером служат исследования т.н. феномена «зловещей долины» [3; 8], в которых зависимость эмпатии к роботу от степени его похожести на чело-

века не являлась монотонной функцией. В таблице 1 представлена схематизация данной ситуации.

В таблице 1 представлена схематизация данной ситуации по результатам классического эксперимента [10]. Табличная, а не графическая, как обычно, форма схематизации использована для того, чтобы легче было понять логику авторского эксперимента. Именно провал оценки (симпатия 0% вместо ожидаемых 80% в столбце D) был назван «зловещей долиной».

Таблица 1

### Схематическое представление результатов экспериментов

Итоговая шкала C (%). Симпатия	Шкала 1 A Степень антропоморфности робота (%)				
	A	B	C	D	E
	20	40	60	80	100
100					100
80					
60			60		
40		40			
20	20			0	

### Гипотеза исследования

Системный анализ работ по теории катастроф позволил предположить, что нарушения монотонности, в том числе наличие «зоны нечувствительности» и конфигураций типа «клюв» или «зловещая долина», можно объяснить одной из моделей этой теории [5]. Для этого требуется найти ещё, как минимум, ещё одну переменную управления, т.е. предложить дополнительную шкалу. Подтверждением гипотезы о наличии катастрофы, т.е. скачкообразных изменений в процессе оценивания, обусловленных гистерезисом, будет являться зависимость результатов от последовательности предъявления стимулов.

### Основная часть

В наших экспериментах к шкале 1 антропоморфности (A) мы добавили шкалу 2 безопасности (B). Обоснованием этого явилось убеждение, что симпатичными и несимпатичными, иначе, привлекательными и непривлекательными, могут быть как люди, так и различные предметы, в том числе неодушевлённые, причём независимо от сходства с человеком. Соответственно, каждый объект имел две исходные характеристики по шкалам A и B и итоговую оценку C.

Для эксперимента были отобраны 25 фотографий роботов, игрушек и персонажей фильмов. Отбор осуществлялся таким образом, чтобы организаторы эксперимента могли классифицировать их по шкалам степени антропоморфности (A) и безопасности (B), а также разместить по одной фотографии в каждой позиции матрицы эксперимента, названной Матрица-25 (табл. 2).

Для удобства внесения изменений в ходе эксперимента координаты ячеек матрицы были взяты аналогичными координатам шахматной доски, а значения, соответствующие шкалам, использовались как дополнения к координатам. В Матрице-25 для каждой ячейки обозначены верхние пределы значений по соответствующей шкале. Каждая фотография изначально соответствует определенной ячейке матрицы. Выбранный нами вариант представлен на рисунке. В ходе эксперимента каждую фотографию можно было заменить фотографией с эквивалентными значениями, однако мы этого не делали.

Таблица 2

**Содержание и категории, использованные в разработке методики диагностики эффекта «зловещей долины» (Матрица-25)**

Шкала 2 <i>Б</i> (%)	Градации (координаты по шкале 2)	Шкала 1 <i>А</i> Степень человекообразности робота, %				
		А	В	С	Д	Е
		20	40	60	80	100
100	<b>1</b>	<i>A1</i>	<i>B1</i>	<i>C1</i>	<i>D1</i>	E1
80	<b>2</b>	<i>A2</i>				E2
60	<b>3</b>	<i>A3</i>				E3
40	<b>4</b>	<i>A4</i>				E4
20	<b>5</b>	<i>A5</i>	<i>B5</i>	<i>C5</i>	<i>D5</i>	E5

В ходе эксперимента все 25 фотографий Матрицы-25 (табл. 2) в определенной последовательности были представлены участникам на большом экране с помощью проектора для оценки по итоговой шкале. Оценки симпатии выставлялись участниками на специальных бумажных бланках по шкале от -5 до +5, где отрицательная оценка соответствовала уровню антипатии к объекту, а положительная – симпатии.

Опрос был анонимным. В эксперименте участвовали школьники и студенты Владивостока и Омска, всего около 120 человек в возрасте от 14 до 35 лет. Каждый эксперимент проводился группой от 7 до 25 человек.

Проверка гипотезы о модели катастрофы в виде сборки Уитни потребовала использования двух вариантов обхода точки бифуркации. В наших обозначениях это соответствовало предъявлению фотографий в направлении от E1 до A1 и далее по шкале А, а в другом – от E1 до E5 и далее по шкале Б. Смысл вариантов был следующим. Левый обход (L): сначала снижается степень сходства с человеком, затем увеличивается угроза. Правый обход (R): сначала увеличивается угроза, затем степень сходства с человеком уменьшается. Для каждого варианта была заготовлена соответствующая презентация.

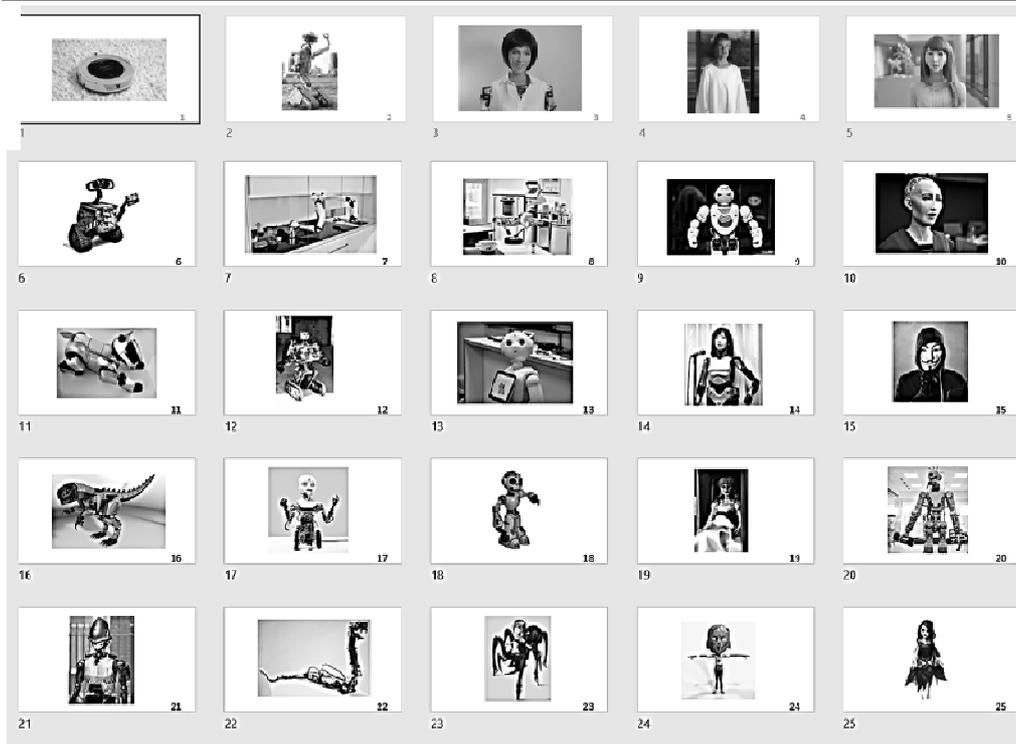


Рисунок. Изображения объектов в соответствии с Матрицей-25 (табл. 2)

Результаты нашего пилотного эксперимента показали существование значительной разницы оценок для разных вариантов обхода. Это даёт основание для разработки более детализированной методики исследования феноменов типа «зловещей долины», учитывающей личностные характеристики участников, например, тревожность и т.п. [9].

#### **Выводы и научная новизна**

Показаны результаты исследования феномена «зловещая долина», а также работ, посвященных анализу этого явления с позиции разных научных направлений. Предложен междисциплинарный исследовательский взгляд на изучение данного феномена, в рамках которого используются системный подход и разработка методики психодиагностики.

Методика эксперимента описана достаточно подробно. Это даёт возможность коллегам реализовать собственные эксперименты по изучению эффектов, связанных с катастрофами восприятия и экспертизы.

Представляется интересным использовать рефлексивные модели В.А. Лефевра, в которых есть параметр, отражающий историю взаимодействия субъекта с определённым стимулом. Однако эти модели являются непрерывными функциями, и их использование требует дополнительных исследо-

ваний. Дальнейшие исследования связаны с созданием и изучением систем коллективной оценки.

1. Ананьина О.А. Протезированные идентичности в романе П. Уоттса «Ложная слепота» // Известия Южного федерального университета. Филологические науки. – 2014. – № 1. – С. 18–27.
2. Климова С.М. Androidscience и неомифологизм // Философские науки. – 2013. – № 9. – С. 44–53.
3. Лефевр В.А. Что такое одушевлённость? – 2-е изд. – Москва: Когито-центр, 2017. – 123 с.
4. Орлова Ю.С. Исследование отношения к культурному интерфейсу социального робота // Проблемы современной психологической теории и практики: сборник научных статей / отв. ред.: Ж.М. Елисеева, Р.В. Осин. – Москва: Перо, 2018. – С. 78–81.
5. Постон Т., Стюарт И. Теория катастроф и ее приложения / пер. с англ. и предисл. А.В. Чернавского. – Москва: Мир, 1980. – 607 с.
6. Особенности проявления эффекта «зловещей долины» у людей разных возрастов / Л.Э. Семенова, А.Р. Сергеев, М.В. Мольков [и др.] // Нижегородский психологический альманах. – 2020. – № 2. – С. 9.
7. Серкин В.П. Методы психологии субъективной семантики и психосемантики: учебное пособие для вузов. – Москва: ПЧЕЛА, 2008. – 382 с.
8. Is the Uncanny Valley an Uncanny Cliff? / C. Bartneck, T. Kanda, H. Ishiguro, N. Hagita // Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN, 2007, 26-29 August, Jeju, Korea. New York: IEEE, 2007. P. 368–373.
9. Ellenberg J. How Not to Be Wrong: The Power of Mathematical Thinking. – New York: Penguin Group, 2015.
10. Masahiro M., MacDorman K., Kageki N. The Uncanny Valley // IEEE Robotics & Automation Magazine, 2012. – P. 98–100.
11. Darintsev O.V., Migranov A.B. A Step-by-Step Algorithm for Finding the Optimal Strategy for the Behavior of a Group of Robots // Mechatronics, Automation, Control. – 2020. – Vol. 21, № 12. – P. 689–695.

#### Транслитерация

1. Anan'ina O.A. Protezirovannye identichnosti v romane P. Uottsa «Lozhnaya slepo-ta» // Izvestiya Yuzhnogo federal'nogo universiteta. Filologicheskie nauki. – 2014. – № 1. – S. 18–27.
2. Klimova S.M. Androidscience i neomifologizm // Filosofskie nauki. – 2013. – № 9. – S. 44–53.
3. Lefevr V.A. Chto takoe odushevlyonnost'? – 2-e izd. – Moskva: Kogito-centr, 2017. – 123 s.
4. Orlova Yu.S. Issledovanie otnosheniya k kul'turnomu interfejsu social'nogo robota // Problemy sovremennoj psihologicheskoy teorii i praktiki: sbornik nauchnykh statej / отв. red.: Zh.M. Eliseeva, R.V. Osin. – Moskva: Pero, 2018. – S. 78–81.
5. Poston T., Styuart I. Teoriya katastrof i ee prilozheniya / per. s angl. i predisl. A.V. Chernavskogo. – Moskva: Mir, 1980. – 607 s.

6. Osobennosti proyavleniya effekta «zloveshchej doliny» u lyudej raznyh vozrastov / L.E. Semenova, A.R. Sergeev, M.V. Mol'kov [i dr.] // Nizhegorodskij psihologicheskij al'manah. – 2020. – № 2. – S. 9.
7. Serkin V.P. Metody psihologii sub"ektivnoj semantiki i psihosemantiki: uchebnoe posobie dlya vuzov. – Moskva: PCHELA, 2008. – 382 s.

© В. А. Филимонов, 2021

© В. С. Чернявская, 2021

**Для цитирования:** Филимонов В. А., Чернявская В. С. Катастрофы субъективных оценок и их модели: к вопросу инструментализации диагностики эффекта «зловещей долины» // Территория новых возможностей. Вестник Владивостокского государственного университета экономики и сервиса. – 2021. – Т. 13, № 2. – С. 170–178.

**For citation:** Filimonov V. A., Chernyavskaya V. S. Disasters of subjective assessments and their models: to the question of instrumentalization of diagnostics of the "uncanny valley" effect, *The Territory of New Opportunities. The Herald of Vladivostok State University of Economics and Service*, 2021, Vol. 13, № 2, pp. 170–178.

DOI <https://doi.org/10.24866/VVSU/2073-3984/2021-2/170-178>

Дата поступления: 17.05.2021.