arXiv:2302.03219v1 [cs.RO] 7 Feb 2023

社交机器人的身体意象*

Bing Li[1], Oumayma Ajjaji[1],Robin Gigandet[1] 和 Tatjana Nazir[1]

摘要——社交机器人的迅速发展，对机器人学和认知科学提出了挑战：如何理解人类对机器人外观的感知。本研究分析了人类自发生成的与机器人相关的词语，以从语义层面揭示过去几十年来研发的 30 种机器人的身体意象。分析利用了词语情感量表和嵌入向量，并为人类感知与身体意象之间的联系提供了一系列证据。研究发现，身体意象的效价和支配性反映了人类对机器人这一一般概念的态度；机器人的用户群体和用途是影响人们对单个机器人印象的主要因素；机器人的情感与其同"人"这一词语之间的语义距离存在关联。根据研究结果，为机器人构建身体意象是一种有效的范式，可用于探究人们欣赏哪些特征，以及哪些因素会影响人们对机器人的感受。

I. 引言

本研究旨在了解人类如何感知机器人，以及如何仅根据视觉特征形成对机器人的语义表征。"身体意象"一词将用于指代一个人通过观看机器人（的图像）所获得的语义理解。

神经科学研究表明，输入大脑的视觉信息会通过一条默认路径转化为语义表征 [1]。笔画、形状和颜色等视觉特征会在枕叶中进行处理，随后通过两条通路传递至额叶和颞叶，以对物体进行语义解释 [2]。重要的是，面孔和词形等显著特征会在靠近枕叶区域的专门区域中进行处理 [3], [4], ，之后才会在颞叶中处理其他语义信息 [5], [6], [7]。最近，有研究提出了第三条通路，专门用于处理社会线索 [8]。因此，机器人的身体意象不仅包括机器人的视觉外观，还包括其语义含义和社会意义。

语义学研究为概念如何在大脑的语义系统中组织提供了模型。在心理语言学中，常用的模型是词嵌入模型 [9]。该模型利用单层神经网络，计算词语在通常可从互联网获取的大量自然生成文本中相邻出现的可能性。

*本项目获得了MEL（里尔欧洲大都会）和I-SITE ULNE（里尔大学北欧）的资助。资助授予TAN（编号：R-Talent-20-006-Nazir）。[1]
作者隶属于SCALAB、UMR CNRS 9193、里尔大学（法国里尔）。如需联系作者，请联系Bing Li（bing.li@univ-lille.fr）或 Tatjana Nazir（tatjana.nazir@univ-lille.fr）

表I 用于测量参与者对机器人这一一般概念的态度的态度问卷。该问卷在自由联想任务之前进行。

说明	请选择最能 描述你此刻态度的选项:
陈述	我希望有一个机器人来协助我..... （1）.....在家中 （2）.....在学校 （3）.....在危险场所 （4）.....在工厂 （5）.....在医院 (6) . . . 在酒店中 (7) . . . 在博物馆中 (8) . . . 在办公室中 (9) . . . 在警察局中 (10) ... 在公共交通工具上 (11) ... 在购物中心中 (12) ... 在体育设施中
回答	□ 非常不同意 □ 不同意 □ 既不同意也不反对 □ 同意 □ 非常同意

研究表明，该模型能够准确反映现实世界中的语义信息 [10]。因此，将人对机器人的功能的理解纳入这样的语义框架，可能有助于改进机器人身体意象的模型。与其关注通常难以量化的机器人的视觉外观，我们因此提出通过分析人们自发联想到机器人的词语来推断机器人的身体意象。

为了了解人类对社交机器人的心理表征，我们提出从语义视角考察机器人的特征。我们将采用自由词语联想任务，收集人们对各种机器人的基本印象，并利用词向量在语义空间中为这些数据建模。我们还将使用情感词典 [11] 中的"效价""唤醒度"和"支配性"指标，对词语的情感维度进行量化。最后，我们将确定表征特定机器人的词向量与语义空间中的"人"一词有何关联，以评估参与者在多大程度上将机器人与人类环境联系起来。综上，本研究表明，表征对于研究人机交互至关重要。

II. 方法

A. 机器人与人类参与者

我们从过去几十年间市场上销售或作为原型出现的机器人中选取了30个机器人。这些机器人

的选择涵盖了广泛而多样的特征，例如机器人是否拥有面部、腿部或手臂，或者是否完全不像常见的动物或人类形状。纳入研究的机器人包括：Buddy、Eilik、Musio、Pico、Sima、Aibo、Ameca、Aquanaut1、Asimo、Astro、Atlas、Emiew、Hexa、Hitchbot、iCub、Jibo、Lovot、Minicheetah、Moxie、Nao、Neubie、Nicobo、Optimus、Parky、Pepper、Pyxel、Sawyer、Spotmini、Talos、Vector。

值得一提的是，机器人是一个含义模糊的概念，有时很难判断哪些对象应当归入其中。一个典型的例子是无人机，例如四旋翼飞行器。有些人会将它们视为机器人（例如，无人机被列为 https://robots.ieee.org/robots/ 上的机器人），但也有人认为它们只是航空器。另一方面，本研究旨在探究机器人身体意象，将其作为人机交互的基础，而这并非无人机的典型用途。因此，本研究未将无人机纳入其中。

实验在 Prolific平台 https://app.prolific.co 上进行，并使用 psychJS[12] 构建。发布实验任务前，我们设置了预筛选条件，仅纳入 18 至 30 岁的英语母语者，以确保年龄范围的同质性；处于这一年龄段的人们很可能伴随着互联网成长，也更愿意接受新技术。

在任务开始前，参与者被告知，他们可以随时退出任务，并会按实际花费的时间获得报酬。该任务平均每位参与者耗时14分钟。报酬约为每小时7.15加元。共有30名参与者被招募参加实验。

表 II 基于全部 30 名参与者和 30 个机器人，在自由联想任务中出现频率最高的 30 个词的频次。

	Word	频次
1	toy	42
2	cute	39
3	友好的	33
4	小的	30
5	机器人	29
6	乐于助人的	28
7	可怕的	25
8	未来	23
9	dog	22
10	令人毛骨悚然的	17
11	fun	16
12	人类	16
13	技术	16
14	儿童	14
15	cool	14
16	奇怪的	14
17	动物	12
18	简单的	12
19	快乐的	11
20	智能	11
21	强烈	11
22	人工智能	10
23	助手	10
24	自动的	10
25	颜色	10
26	帮手	10
27	现代的	10
28	有用	10
29	轮子	10
30	自动化的	9

B. 机器人态度量表

实验由两个环节组成：一个环节测量参与者对机器人的态度，另一个环节构建机器人的身体形象。为了测量态度，我们设计了一份问卷。为了构建机器人的身体形象，我们采用了自由联想任务 [13], [14]。参与者需要先完成问卷，然后完成自由联想任务。

机器人态度问卷采用李克特量表，呈现一些陈述，询问受试者是否愿意在不同情境中接受机器人的协助，如表 I 所示。该问卷改编自另一项研究中使用的量表，该研究考察了不同文化之间的差异 [15]。参与者在 5 点李克特量表上选择最能描述其态度的选项：从 1 分“非常不同意”到 5 分“非常同意”。回答被赋予 0 至 4 的分值，分别代表从消极到积极的态度。每位参与者的总体态度得分为所有陈述得分的平均值。该问卷在自由联想任务之前进行，因此 机器人不应代表任何特定的机器人模型，而只代表机器人这一一般概念。

C. 自由联想

问卷结束后，我们请参与者针对依次呈现的 10 张图像，每张写下 6 个脑海中浮现的词语。每张图像都会显示在屏幕顶部，下方则有 6 个文本框。参与者需要按顺序从第 1 个文本框到第 6 个文本框输入词语，然后点击“完成”按钮，继续查看下一张图像。输入一个词语后，他们可以按“Enter”键快速跳到下一个词语。每张图像都必须填写全部 6 个词语，不能跳过任何文本框。我们停用了 psychJS 程序的“自动填充”功能。

为了避免参与者内部相关性造成混淆，即同一参与者对所观察到的不同机器人表现出相似反应，我们采用了机器人随机选择程序。随机化方法已在另一项研究中说明 [16]。通过这一程序，每位参与者都会看到不同的 10 个机器人子集。

我们在表 II 中列出了出现频率最高的 30 个词。

D. 词向量与情感测量

为了用向量表示词语，我们从 FastText 基于维基百科新闻预训练的 Word2vec 模型中获取词向量：https://fasttext.cc/ docs/en/english-vectors.html。这些向量有 300 个维度。由于一个机器人可能对应多个词，因此我们对各维度的向量取平均值，生成一个代表该机器人的 300 维向量，如图 2 所示。

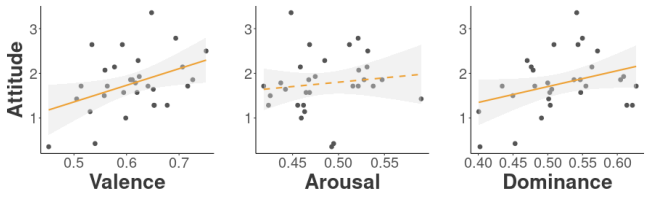


图 1. 态度与支配性、效价和唤醒度之间的关系。虚线表示相应 LME 分析中的结果不显著。

为了量化词语的情感维度，我们采用 Mohammad 的《情感词典》中针对 20,000 个英语单词、使用最佳-最差标度法得出的效价、唤醒度和支配性（VAD）测量值（2018年） [16]，[11]。收集数据后，我们对任务中的回答进行了预处理，将复数词转换为单数形式，并将大写词转换为小写词。对于复合词和短语，我们主要提取其中最具代表性的词，例如，将“canfly”替换为“fly”，将“car-like”替换为“car”；若难以判断哪个词更具代表性，例如“artificial intelligence”，则舍弃整个复合词。有些形容词未收录在该词典中，但其同义词有收录，例如，“gangly”未收录，而“awkward”有收录。在这种情况下，我们根据《韦氏同义词词典》（https://www. merriam-webster.com/thesaurus/）将其替换为同义词。预处理后，效价-唤醒度-支配性词典覆盖了 92.33% 的回答。在该词典中，情感维度的取值范围为 0 到 1，数值越大表示情感评分越高。

III. 结果

A. 词语情感预测主观态度

在参与者层面，我们考察了能否根据关联词的情感测量指标预测其对机器人的总体态度。我们建立了三个线性混合效应（LME）模型，分别以态度评分为因变量，并各自仅将情感测量指标作为固定效应自变量。将机器人作为随机截距纳入模型，以构建最大化随机效应结构。

结果显示，效价是态度的显著预测因子（似然比检验，LRT $\chi^2 = 15.577$，$p < .000$）；支配性也是态度的显著预测因子（LRT χ^2 =19.529，$p < .000$）；而唤醒度无法预测态度（LRT $\chi^2 = 2.101$，$p = .146$）。

态度与情感之间的关系如图 1 所示。情感值是对每位参与者对应的词语取平均后得到的，也就是说，每个点代表一位参与者。总体而言，结果验证了机器人身体意象的语义联想反映了观察者的态度。

B. 身体形象图

人们通常会从多种角度对机器人进行分类，例如机器人的形态、功能和用途等。然而，这些分类角度之间的相对位置，或哪种角度最受受试者偏好，几乎尚未得到确定。

在这里，我们根据机器人的联想词向量构建了一个机器人图。我们对每个机器人的向量取平均值，以表示该机器人。在图中，每个机器人都与另外三个与其最相似的机器人相连。根据向量表示和两两之间的余弦相似度，这些相连的机器人构成了一个包含30个机器人的图，如图2和图3D所示。在身体图中，我们识别出一些通常被称为人形机器人、机器人助手、机器宠物等的机器人类型。然而，有些机器人无法明确归入某一类别。例如，Aibo 常被称为“机器狗”，但它与 Eilik 的连接比与 Spot 更紧密；而 Neubie 的外形像手推车或机车，似乎更可能与人形机器人相连，而不是与外形相似的 Parky 及其他轮式机器人相连，例如 Pico 和 Astro。

尽管如此，通过找出团，即图中完全连接的组件，仍然可以识别出最具区分度的类别。我们在图中用下划线标出了4-团，以说明机器人类别，即（A）机器宠物、（B）人形机器人和（C）机器人儿童（见图3）。Nao 似乎完美地属于面向儿童的机器人类别，其中包括机器宠物和机器人儿童。看到 Nao 的人会立即联想到它应当适用的用户群体和用途。相比之下，Neubie 的身体意象不明确，因此更难与特定的用户群体或用途建立联系。总体而言，根据语义关系所识别出的机器人类别表明，受试者倾向于根据机器人预期的用户群体和用途来构建身体意象，而不是基于其形状。

C. 人类距离预测身体意象的情感

身体意象与一个与人类相关的词语之间的距离，可能有助于揭示机器人在语义上与人类生活的接近程度。在这里，我们获得了机器人的平均向量，并计算了它们与“人”一词之间的距离，即 $1 -$ 相似度$_{\cos}$（v_1,v_2），其中 v_1 和v_2 分别是抽象表示机器人和目标词“人”的向量。

我们选择“人”作为代表真实人类生活的目标词，是为了尽可能多地纳入与每个机器人和目标词之间距离相近的词语，用以控制针对目标词的情感偏差。此类偏差的一个例子是词语的效价，它与词语到“人类”的距离呈正相关。因此，词语的情感水平往往与其到目标词的距离呈现系统性关系，例如到“人类”的距离。因此，我们从机器人的情感值中减去一个基线，以标准化机器人的情感水平。基线是与机器人距离相近的词语的平均情感水平。具体而言，在语义空间中，我们计算了机器人与目标词之间距离差的平均值，并以此作为一个掩码，其中心位于机器人到目标词的距离处。我们提取情感词典中被掩码覆盖的词语，并

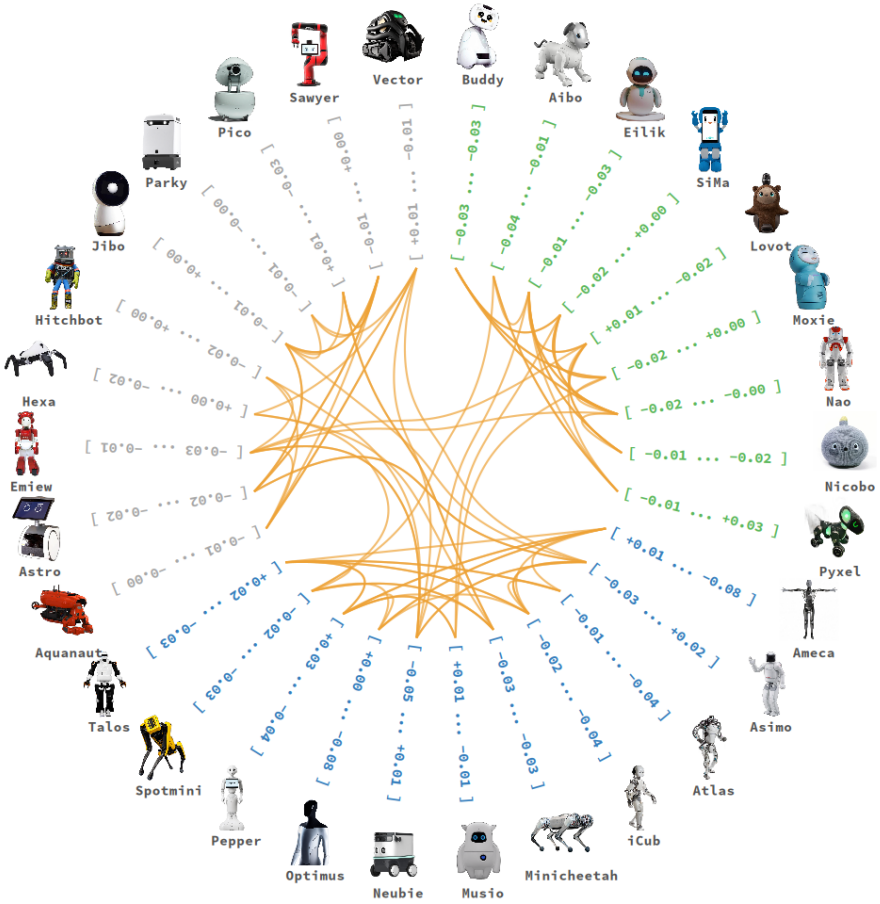


图2。基于余弦相似度对机器人进行聚类。最外层圆表示本研究中的机器人。中间圆表示机器人的向量抽象，并根据机器人之间余弦相似度的层次聚类结果，以三个聚类的颜色进行标示。最内层网络由每个机器人与其3个最近邻之间的连接构成。

将其用作情感基线词。以“人类”和“人们”为中心的基线词数量远少于或少于以“人”为中心的基线词。因此，以“人”作为目标词能够提供最稳健的基线。

在此，我们通过考察人与机器人的距离与身体意象情感之间的关系，研究人们感知到的情感。图4分别展示了效价、唤醒度和支配性的结果。我们发现，效价和支配性都是人与机器人的距离的函数。效价起初随着人与机器人的距离增加而上升，随后在高于基线的某一水平处趋于平稳并略有下降；支配性则随人与机器人的距离单调增长。另一方面，尚未发现人与机器人的距离与唤醒度之间存在简单的关系。

有趣的是，我们发现大多数人形机器人在人与机器人的距离上都比其他机器人更远。例如，距离最远的机器人是 Atlas、Optimus、Pepper、Musio、Hitchbot、Talos、iCub 和 Asimo；距离最近的机器人是 Aquanaut、Jibo、Spot、Aibo、Pico、Pyxel 和 Buddy。结果表明，对于机器人而言，更像人类未必就能更好地融入人类生活。

IV. 讨论

心理语言学提供了关于词语的丰富知识。在此，我们介绍一种范式，使人们能够利用这些知识来分析人机交互。

具体而言，我们分析了与机器人的视觉形象相关联的词语，以探究人类对机器人的感知，并证明情感一词会受到人们对机器人的态度影响。其次，根据相关联的词语，可以按照目标用户群体和外观对机器人进行分组。第三项结果凸显了机器人身体意象的重要性，以及其对人机交互的影响。本研究最后强调，身体形象有助于研究人员和设计师理解机器人的各种维度及其对人类反应的影响，因此具有重要意义。

我们的范式包含两个程序。在第一个程序中，我们将机器人与一系列词语联系起来，从而推断感知情绪，即效价程度（“积极/愉悦”或“消极/不悦”）和支配性程度（“强大/有力”或“无力/弱小”）。图 1 展示了如何利用与词语相关的效价和

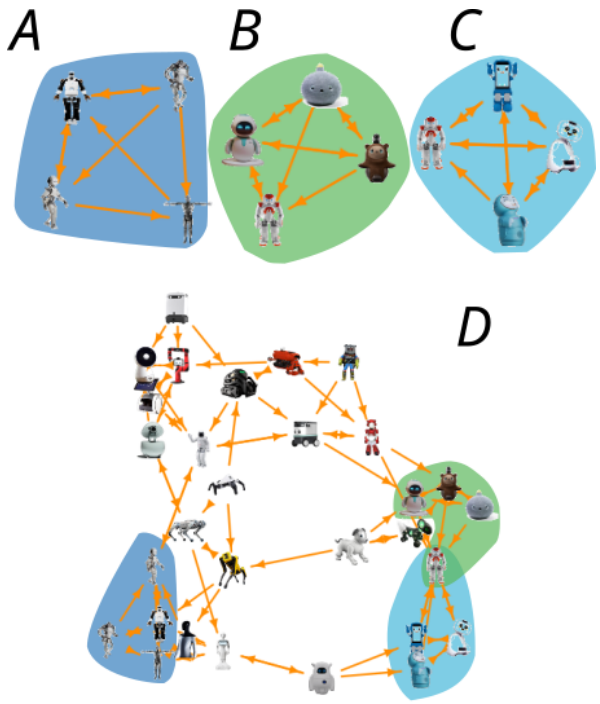


图3. 基于机器人身体意象构建的网络。(A-C) 网络中的三个4-团。(D) 全部30个机器人的整体网络。团的颜色与网络中的颜色相对应。

支配性，根据李克特量表预测人类对机器人的态度。近期另一项关于社交机器人的研究也采用了自由联想 [17]。在该研究中，作者考察了参与者对书面词语“机器人”的联想，并利用因子分析将机器人作为社会实体的心理表征进行分类。我们的研究与这项先前研究具有相同的理论基础，但更具体地聚焦于单个机器人，而非一般的机器人概念或其社会临场感。

在第二个步骤中，词语被转换为向量，从而将机器人抽象为嵌入稳健语义空间中的300维向量。如图3所示，尽管不同机器人的分组是基于与它们相关联的词语，但同一组的成员在外貌及其衍生特征方面表现出相似性。例如，人形机器人聚集在左侧的蓝色团周围，紧邻四足机器人；而面向儿童的机器人则呈现为绿色和浅蓝色的团。增加机器人的数量（本研究仅纳入了30个机器人）后，这种模式可能会更加明显。因此，这些结果支持这样一种观点：机器人的身体意象涵盖相互关联的视觉、语义以及潜在的社会特征，而这些特征均是根据机器人的形态推断得出的。

该范式为人类受试者所感知的机器人提供了信息丰富的描述；通过探查词汇量表和语义空间，这种描述有望变得日益立体。例如，这些机器人彼此之间如何关联，甚至与其中的其他概念如何关联

语义空间中的其他概念（例如“人”这一概念）。在图4中，我们展示了效价和支配性与机器人接近“人”这一概念的程度之间的关系。

最后，发现身体意象可由效价和支配性预测，也可能为商用机器人开发提供启示。例如，图 4 右图所示的分析表明，人形机器人一方面被感知为更具支配性，另一方面则被认为与“人”的距离更远。这表明，人们购买 Jibo 或 Buddy 这类机器人（两者支配性较低，且更接近“人”）的可能性，或许高于购买人形机器人的可能性。

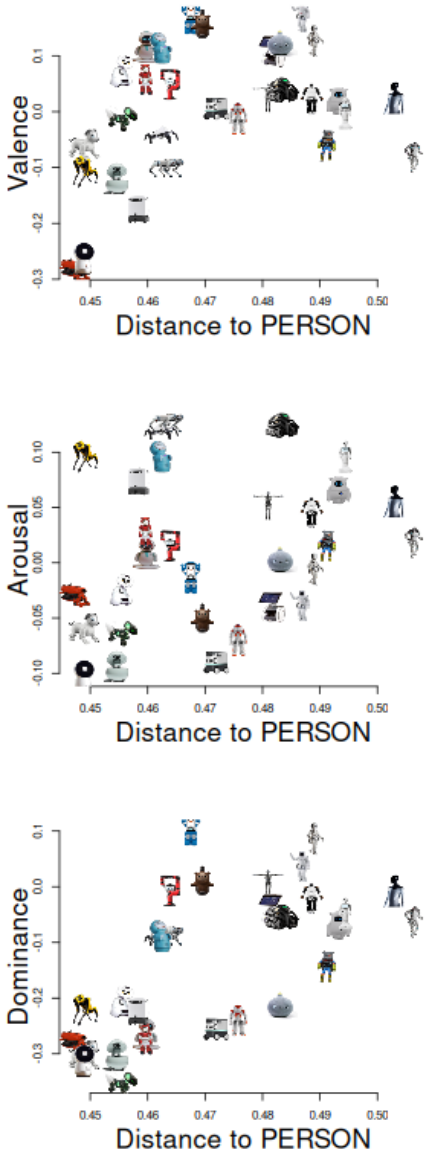


图 4. 情感与人与机器人的距离之间的关系。x 轴表示机器人身体形象与“人”一词之间的距离；y 轴表示标准化的效价、唤醒度和支配性。

参考文献

[1] A. Doerig、T. C. Kietzmann、E. Allen、Y. Wu、T. Naselaris、K. Kay 和 I. Charest，《语义场景描述作为人类视觉的目标》，2022 [ ] 2209 11737 。在线。[2] 可获取于：https://arxiv.org/abs/。K. Kar 和 J. J. DiCarlo，《灵长类腹侧通路需要通过腹外侧前额叶皮层进行快速循环处理，以实现稳健的核心视觉物体识别》，《神经元》，第 109 卷，176，2021 1，第 164no. 页–。[3] T. Hannagan、A. Agrawal、L. Cohen 和 S. Dehaene，《书写词语组合式神经编码的涌现：将卷积神经网络重新用于阅读》，美国国家科学院院刊，第 118 卷，第 46 期，2021。[4] D. Y. Tsao、W. A. Freiwald、T. A. Knutsen、J. B. Mandeville 和 R. B. Tootell，《猕猴大脑皮层中的面孔与物体》，《自然-神经科学》，第 6 卷，第 9 期，第 989–995 页，2003。[5] K. Patterson、P. J. Nestor 和 T. T. Rogers，《你从哪里知道自己所知道的？人脑中的语义知识表征》，《自然综述：神经科学》，第 8 卷，第 12 期，第 976–987 页，12 月。 2007 [ ] 2277 。[6] 在线。可获取于：https://www.nature.com/articles/nrn S. Bracci 和 H. O. de Beeck，《两条视觉通路中形状与类别表征之间的分离与关联》，《神经科学杂志》，第 36 卷，第 2 期，第 432–444 页，2016。[7] J.-Q. Tong、J. R. Binder、C. J. Humphries、S. Mazurchuk、L. L. Conant 和 L. Fernandino，《用于概念多模态经验表征的分布式网络》，《神经科学杂志》，2022。[8] D. Pitcher 和 L. G. Ungerleider，《社会知觉专属第三条视觉通路的证据》，《认知科学趋势》，第 25 卷，第 2 期，第 100–110 页，2021。

[9] S. Bhatia，“联想判断与向量空间语义学。”心理学评论，第 124 卷，第 1 期，第 1 页，2017 年。[10] G. Grand、I. A. Blank、F. Pereira 和 E. Fedorenko，“语义投影从词嵌入中还原了人类对多种物体特征的丰富知识”，《自然·人类行为》，第 1–13 页，2022 年。[11] S. Mohammad，“获取人类对 20,000 个英语单词的效价、唤醒度和支配性进行的可靠评分”，载于计算语言学协会第 *56* 届年会论文集（第 *1* 卷：长篇论文），2018 年，第 174–184 页。[12] J. R. De Leeuw，“jspsych：用于在网页浏览器中创建行为实验的 JavaScript 库”，行为研究方法，第 47 卷，第 1 期，第 1–12 页，2015 年。[13] D. L. Nelson、C. L. McEvoy 和 S. Dennis，“什么是自由联想？它测量的是什么？”《记忆与认知》，第 28 卷，第 6 期，第 887–899 页，2000 年。[14] S. De Deyne、D. J. Navarro 和 G. Storms，“利用连续词语联想而非单词联想所构建的网络，更好地解释词汇与语义认知”，行为研究方法，第 45 卷，第 2 期，第 480–498 页，2013 年 6 月。 [在线]。获取地址：https://link.springer.com/article/ 10 3758 13428 012 0260 7 。/s - - - H. R. [15] Lee 和 S. Sabanovic，“韩国、土耳其和美国在机器人设计与使用方面存在不同的文化偏好”，载于*2014* 年 *ACM/IEEE* 人机交互国际会议论文集，2014 年，第 17–24 页。[16] G. Hollis，“在不平衡的多项目设计中对最佳-最差数据进行评分，并将其应用于众包语义判断”，行为研究方法，第 50 卷，第 2 期，第 711–729 页，2018 年。[17] S. Brondi、M. Pivetti、S. Di Battista 和 M. Sarrica，“我们对机器人有何期待？日常生活中的社会表征、态度以及对机器人的评价”，《社会中的技术》，第 66 卷，文章编号 101663，2021 年。