arXiv:2304.02861v1 [cs.HC] 6 Apr 2023

创建公开人类用户 Video Games 数据集的可重复性与透明度

艾玛·J·普雷蒂* 雷南·瓜雷塞† 海瑟姆·M·法耶克‡ 法比奥·赞贝塔§

澳大利亚皇家墨尔本理工大学

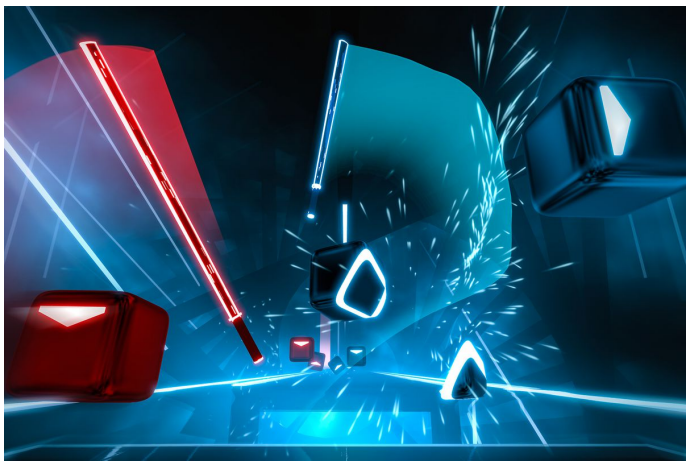


图 1：Beat Games 推出的虚拟现实游戏 Beat Saber [3]

摘要

游戏研究缺乏可重复性；方案细节不透明，阻碍了科学共识的形成，也降低了公开数据集的发表意愿，从而影响这些技术在电子游戏研究中的应用。为此，我们提出并举例说明一系列实验考量因素的用法，例如游戏和材料的选择。本文倡导在发布数据集时交流研究方案，从而帮助研究人员设计实验。

索引词： 以人为中心的计算—协作与社会计算—协作与社会计算系统和工具—开源软件；应用计算—其他领域的计算机—个人计算机和 PC 应用— 电脑游戏

1 引言

与《科学》的许多领域一样，可重复性是人类用户电子游戏研究中的一个重大问题。除了研究人员在相同或相似条件下再次进行相同的实验（称为重复性 [10]）之外，很少有研究会从另一项实验中探索完全相同的实验任务、测量工具和分析，这是系统地佐证某个游戏元素或构念的发现以形成科学共识的一种有力方法。

请考虑以下假设情景：两名独立的研究人员探究睡眠剥夺3小时对视频游戏游玩期间自我报告的认知负荷。第一位研究者使用商用虚拟现实游戏 BeatSaber [3]（虚拟现实（VR）；见图1）以及一份关于认知负荷的定制问卷，而第二位研究者决定使用一款定制的虚拟现实节奏游戏，并通过NASA-TLX 任务负荷指数（NASA-TLX） [16]测量认知负荷。所有其他实验方法和细节均相同。

根据定义，这是概念性重复的一个例子，因为第二位研究人员旨在探究与第一位研究人员相同的潜在概念或假设，但采用了不同的实验设计 [10]。然而，这并不能保证得出相同的结果。可能有许多因素会对此产生影响，但本文关注的重点是使用定制任务和工具来检验这类假设。定制工具本身并无不妥，但往往容易受到实验者偏差的影响，而且其源代码或原始材料通常不会公开发表，供其他研究人员查看或使用 [23]。假设在我们的例子中，第一项研究发现自我报告的认知负荷显著较高，而第二项研究的结果却不显著，我们该如何解释这种差异?

除了可能影响这些结果的更直接变量，例如自我报告偏差、对游戏的先验知识以及睡眠剥夺对个体的影响外，我们认为，测量工具和实验任务本身也应接受更为严格的审视，尤其是在尝试重复一项研究时。与此相关的是透明度；在本例中，这包括对所采用的方法论和程序进行完整记录与传播、提供定制工具或测量工具的源代码，以及提供用于处理和分析数据的代码，这也与可复现性有关 [11]。可重复性与透明度密不可分；缺乏透明度，就不可能对研究进行准确的直接重复或系统性重复，而不成功的重复研究往往会缺乏透明度，或者根本不会发表 [28]。这不仅是科学方法的核心支柱，也关系到公开数据集的创建，而在人类用户电子游戏研究中，这类数据集仅有少数几个。

*电子邮件: emma.pretty@rmit.edu.au
†电子邮件：renan.guarese@rmit.edu.au
‡电子邮箱：haytham.fayek@ieee.org §电子邮箱：fabio.zambetta@rmit.edu.au

在本文中，我们提出了开发能够促进数据集发布的实验设计时需要考虑的一些核心事项：

- 我们说明了选择用于测量感兴趣的变量的工具时需要考虑的因素（第 3 节）
- 我们提供了如何利用游戏作为实验刺激物的详细信息，以促进透明度和可重复性（第 4 节）
- 还列出了一些关于方法论的考虑事项，以尽量减少研究内部和研究之间的变异性，从而加强未来的重复研究（第 5 节）
- 最后，论文以一个示例作结，展示所讨论的理论考量如何在实践中应用，并为所作出的每项选择提供明确的理由（第6节）。

2 相关工作

2.1 作为刺激物的游戏

许多研究都阐述了如何选择和使用电子游戏作为实验任务的一些关键标准，并指出了使用现成游戏和定制游戏的优点与局限 [17, 23]。人们普遍认为，现成游戏虽然能带来更好的生态效度（研究的可推广性），但其刺激物较为复杂，可能难以就游戏过程的某个单一方面得出恰当的结论。由于现成游戏通常并非开源，因此也往往不易修改。另一方面，定制游戏不够精致、吸引人和自然，这不利于外部效度，也不利于对玩家行为作出判断。不过，定制游戏确实有助于研究者完全控制实验环境。定制游戏的另一项主要局限是实验者偏差——在创建游戏时，研究者已知悉正在调查的构念，因此实验设计可能会（有意或无意地）倾向于产生符合预期的结果。

超越传统的电脑或主机游戏，同样的考量也适用于扩展现实游戏。例如，Yildrim [36]选择了两款现成游戏来比较其晕动症，依据是这两款游戏都具有可比的桌面版和虚拟现实版本；在两个版本中，玩家视角都可以通过控制器输入或头部运动进行改变；并且，两款游戏在快速推进的环境中导航时，都会产生视觉诱导的自运动。类似地，Shelstad 等人 [31] 使用了一款具有桌面版和虚拟现实版本的现成塔防游戏，在两个平台上通过游戏满意度量表评估用户体验。在这两项研究中，研究人员选择了已经确立的商业游戏，因为它们的特征与研究需求相符。另一方面，学术界已经针对特定研究目的提出、开发并进行用户测试了几款定制的虚拟现实游戏，通常处于教育或严肃游戏的情境中 [8, 29]。然而，就可复现性和透明度而言，这两种选择都是合理的，但在考虑发布研究所得的数据集时，确实需要额外考量；我们将在本文后文对此进行介绍。

2.2 免费获取数据

许多综述和文章都呼吁，必须提供作为科学出版物基础的数据，这也是科学方法的根本 [1]。《科学》建立在根据数据得出结论的基础之上，而且将近一半的研究人员会在某个时候使用其他科学家生成的数据 [32]。尽管如此，所有科学领域的研究人员都不愿分享数据，也不会提供他们的数据。例如，在一项针对488名经济学研究人员的实证研究中，研究发现，80%的研究人员不会主动分享他们的数据 [1]。另一项研究 [32]请求获取310篇论文的数据，这些论文都包含数据可用性声明（例如“联系作者获取数据”），并发表于《自然》或《科学》。研究发现，只有39.4%的作者接受了请求，19.4%的作者拒绝了请求，41.3%的作者没有回应（60天最长响应期限）。尽管大多数人拒绝并忽略了这些请求，但研究发现，带有代码库数据链接的论文，其引用次数最多可增加25% [9]。虽然将数据与其对应的出版物一同发布总体上有益于科学，但这些数据并不总是可用。这可能是由于出版商的限制或数据的保密要求所致，但一个可自由访问的数据集面临的主要障碍，是糟糕的注释和文档导致其完整性与可理解性不足 [27]。机器学习领域已经尝试对已发布的数据集进行标准化，以便用于应用人工智能算法。[14] 的作者提出了一系列涵盖整个科学流程的问题，供研究人员在认真反思和设计数据集及其配套文档时使用。同样，NeurIPS会议也为其作者提供了一份指南清单 [13]，其中涉及透明度、可复现性和伦理。这份清单已被用于评估研究传播，旨在激励研究人员认真考虑这些方面。

具体而言，关于虚拟现实用户研究，Lanier 等人 [21] 对涉及数据分析透明度和有效性的61篇学术论文进行了元分析。他们的研究揭示了统计报告中较高的错误率，以及所调查研究总体上较低的透明度，其中公开提供数据集的研究不到4%。

2.3 人类用户数据集

关于视频游戏数据集，大量数据集旨在应用人工智能算法（参见 [19]，了解一些常见的游戏人工智能平台和数据集概览），通常由公开可用的游戏过程视频片段组成，例如 [22] 和 [15]。然而，使用在游戏过程中实时收集的人类用户数据的数据集却很少。我们将给出一些近期值得注意的数据集示例。

AGAIN 数据集 [24] 包含 37 小时的游戏过程录像，涵盖 9 款专为该数据集创建的不同游戏。这 9 款定制游戏都深受现成游戏的影响。尽管它们是为该数据集而创建的，但目的并非唤起某一种特定情绪，而是观察玩家的情感反应。该数据集有详尽的文档说明，提交使用条款协议后即可轻松获取。FUN*ii* 数据库 [4] 包含 190 名玩家在游玩育碧《刺客信条：大革命》 [33] 和《刺客信条：枭雄》 [34]期间采集的生理数据。作为游戏领域中同类首个数据集，它包含心电图、皮肤电活动和肌电图等数据类型。相关出版物有详尽的文档说明，但由于链接失效，撰写本文时尚无法获取该数据集。平台游戏体验数据集 [20] 采用了原版《马里奥兄弟》的仿制版本，主要包含行为数据，详细记录参与者的行为、游戏偏好和人口统计信息；此外，还包括经过预处理的游戏过程中参与者面部特征数据。该数据集可通过 Google Drive 文件夹链接免费获取。

特别关注虚拟现实研究时，可以发现该领域在数据分析及其可用性方面的透明度明显较低，其中一个例子在2.2 节中有所讨论。[21] 在许多情况下，即使研究者努力发布所使用的数据，所发布的数据通常也仅限于“支持本研究发现”的数据 [2, 36], ，而没有明显的途径访问所收集的全部数据以进行进一步研究。

3 测量

除了确保研究的构念效度之外，选择用于测量目标变量的工具也关系到研究的可重复性。为了让其他研究人员也能使用该数据库，这些工具应与研究相关，同时也应提供这些工具的源材料。也就是说，如果采用了多份自我报告问卷，那么它们.....

如果研究人员能够访问所使用的问题，将会从中受益。这既适用于研究中采用的现有测量工具，也适用于研究中实施的任何新问题或问卷。从数据集创建的角度来看，这可能意味着同时纳入问卷中的原始数据，以及对问卷子集的任何综合结果，并在研究成果发布时，或在数据存储库中，附上这些问题（以及评分说明）。问卷通常会根据其他领域的问卷进行调整，然而，如果不直接联系研究人员，便无法获取这些经过调整的问题。这些内容也应在工作成果发布时提供；如果允许提供的详细程度有限，则应提供一个指向永久存储库的链接，其中包含其他补充材料。

对于用于测量目标变量的硬件（例如情感计算研究中经常使用的心率传感器），由于成本和可用性方面的限制，在设备选择上可能缺乏灵活性或自由度。某些生理硬件在质量和可获取性方面通常存在差异（例如 NeuroSky 的 Mindwave[1] 与 Brain Product 的 ActiCHamp[2]相比），不过仍可考虑这些设备的某些特性。例如，有些设备及其配套软件完全采用开源形式，使研究人员能够完全了解设备的内部运作机制。研究人员更倾向于使用这些设备，因为通过重新分享对实验所做的任何修改（模组）或添加内容，它们能够促进研究人员之间的便捷协作。近期的一些例子包括 OpenBCI 的 Ultracortex 脑电图（EEG）头戴设备[3] 和 Emotibit[4]，它们可以测量多种生理数据（心率、皮肤电活动等）；此外，还有 DeepFace[5] 和 OpenFace[6]等开源面部识别软件。为了协助研究人员，我们在此处以及后文提供了一些问题，供研究人员在设计实验时思考，以便考虑透明度和可重复性实践：

- 如果使用成熟的测量方法，它是否曾在此情境中经过验证或使用过?
- 研究人员必须付费才能获取这项测量数据吗？还是可以低成本获取?
- 我可以分享这个工具（例如附录中的完整问卷问题集），还是有其他更易于获取的内容可供我使用?
- 我正在制作自己的问卷，如何确保读者能够获取完整的问题集、评分信息和设备信息?

4 游戏选择

关于游戏选择，适用的原则与测量选择类似，但还需考虑一些额外因素。在为研究选择游戏时，首先必须决定是使用现成游戏，还是为研究目的创建定制游戏。这两种选择各有优点和缺点，而且都无法完全避免透明度方面的问题。

定制。 正如 [17], 研究人员创建的定制游戏所强调的那样，这些游戏具有诸多优势，包括能够控制所有实验细节和潜在的混杂变量。然而，定制游戏也会引入一系列偏差问题，因为它们是专门为解决相关研究问题而设计的。定制游戏通常不会公开发布，这限制了其他研究人员复现同一研究的能力。这不仅阻碍了基于该实验设计开展进一步工作，也降低了实际任务本身的透明度。通常会提供文字描述，但这些描述不一定足够充分。因此，读者可能无法察觉任务设计或实现中存在的偏差，而这些偏差可能会使研究结果倾向于预期方向。因此，我们建议，在研究中选择或需要定制游戏时，应将源代码公开在代码库中，既能让实验的细节更加透明，也能方便研究人员复现该研究。

现成游戏。 除了基于选择游戏的有效性目的之外，使用现成游戏还有其他考量，这些考量可能有助于或阻碍透明度和可重复性。现成游戏是很好的选择，既有利于研究发现的普遍性，也能视游戏而定，允许修改实验环境以实现更好的控制。在某些情况下，游戏拥有强大的修改系统（通常称为模组制作），几乎可以更改、分析或记录游戏的所有方面。例如，Minecraft [26] 的模组制作 API 自其测试版发布以来一直在开发和维护，而 Creation Kit [6] 则支持修改《上古卷轴》系列中的大多数游戏，以及 Bethesda 制作的许多其他游戏。购买这些游戏后，用于修改游戏的软件可免费使用，因此研究人员能够使用现成游戏开展研究。

与定制游戏类似，研究人员可能无法通过足够详尽的文字描述来支持游戏的复现。不过，只要现成游戏可供公众访问，其基本信息通常已经可以获取。话虽如此，在实验设计中对现成游戏所做的任何修改，都必须在描述实验流程时明确说明。此外，也有一些平台允许人们上传对游戏所做的修改；研究人员应利用这些平台，以便轻松分发修改内容，并提高实验内部运作的透明度。例如，Minecraft [26], 主要使用 CurseForge[7]（玩家也可通过该平台下载并应用这些修改，以增强游戏过程）；而《上古卷轴 V：天际》等 Bethesda 游戏则使用 NexusMods[8]。使用支持修改的游戏，其一项核心局限在于可供使用的游戏数量和类型会因此减少；不过，透明度和可重复性的提升所带来的收益足以弥补这一局限。此外，用户生成内容和游戏修改正日益流行，越来越多的新游戏支持添加或修改游戏内容；Garry's Mod [12] 和 Roblox [30] 就是这类游戏的例子，玩家可以使用 Lua 编程语言在其中创作自己的游戏。总之，在实验设计过程中，研究人员应考虑以下一些问题：

- 我正在使用一款现成游戏，可以把我对游戏所做的修改上传到哪里，供他人访问?
- 我正在使用一款定制游戏，能否将其公开发布到代码库中？如果不能，有没有可以使用并公开发布的游戏，或者已经公开发布的游戏?

5 程序

5.1 方法论考量

本节将详细介绍一些最佳实践方法的示例，以防止原始研究与复现研究之间出现差异。其中并非所有方法都适用于每款游戏，但在实施游戏任务和事件编码时，仍建议加以考虑。

无论是哪种游戏，在大多数游戏研究中，都需要向玩家说明如何与任务互动或控制角色。为确保参与者之间的一致性，这些说明需要预先编写成脚本，因为不应让任何一位参与者比另一位参与者获得更多或更少的任务信息。

[1]https://store.neurosky.com/pages/mindwave [2] https://www.brainproducts.com/solutions/actichamp/ [3] https://shop.openbci.com/products/ultracortex-mark-iv[4] https://www.emotibit.com [5] https://github.com/serengil/deepface [6] https://github.com/TadasBaltrusaitis/OpenFace

[7]https://www.curseforge.com/ [8] https://www.nexusmods.com/

然而，由研究者口述的脚本仍容易受到人为错误的影响，研究者的声音相关因素也可能影响参与者的理解。因此，我们建议使用屏幕文字（这在人类用户研究中通常也很常见 [35]）；或者，为了增强游戏的沉浸感，由游戏中的角色或旁白提供音频或文字说明（最好两者兼有），例如使用文字转语音算法生成旁白。对于无法直接整合这些内容的游戏，也可以使用伪旁白，将语音播放在游戏画面之上。整合语音的音频文件或源代码也应在实验详情和数据的文档及发布材料中提供。提供说明脚本还将便于开展其他语言的重复研究，因为翻译时无需依赖研究者转录口述音频。

关于日志记录和事件编码，游戏事件的编码必须精准，并且能够与游戏过程中收集的其他数据进行校准。对于支持修改的商业游戏，这类信息通常很容易获取，因此它是事件编码的黄金标准，也便于开展重复研究。研究人员创建定制游戏时，对游戏的全面掌控可确保游戏中包含功能完善的日志记录器。与其他数据和方法一样，也应提供日志记录器的源代码，以便进行数据同步和满足技术需求。对于无法访问原生或内置日志记录系统的游戏，也有其他替代方案。正如 [18], 所建议的，可以记录按键和鼠标点击，同时进行带时间戳的屏幕录制，以便后期处理。但这并不理想，因为这种方法既耗时，如果需要人工根据屏幕录制标注行为，还容易出现人为错误；这对于精准的数据分析而言尤其成问题。

内部效度直接关系到一项研究能否被复制 [25]。在使用软件作为刺激物的游戏及其他数字研究中，玩家可能会通过作弊或破坏实验来影响这种内部效度。他们可以通过访问特定设置，或杀死、摧毁特定的非玩家角色或物品来做到这一点。为了尽量减少这种情况，需要移除玩家作弊的能力。这在现成游戏中很大程度上是一个问题，因为在游戏过程中，玩家原本可能需要访问特定设置，但这对于实验环境而言并不理想。根据游戏的不同，可以在开发模组时于内部禁用这些设置，有时也可以仅通过创建存档文件来实现，Minecraft [26]就是这种情况。在某些情况下，玩家的角色死亡也可能会通过在其他地点、实验区域之外重生角色而干扰实验。这并不适用于所有研究，因为角色死亡是许多游戏的自然组成部分；但是，在可能的情况下，如果角色死亡并非研究所必需，建议阻止角色死亡。

5.2 程序文档

可重复性研究以能够获取原始研究的实验细节为前提 [11], ；因此，为了促进可重复性，研究人员应尽可能纳入获准提供的全部信息和源代码。正如前文所讨论的示例所示，这包括各个方面的详细信息，从计算机规格到向参与者提出的确切调查问题。在创建公开数据集时，这一点尤其重要，有助于研究人员充分了解分析数据集和应用算法时可能存在的混杂变量。同样，建议纳入所有测量过的数据和材料，即使它们并非研究重点；这有助于提高透明度，确保纳入不利于原始研究问题或假设的研究结果。然而，必须指出的是，重要的是要出于保护参与者隐私的考虑，应从所提供的数据集中删除所有个人身份信息。还应在数据集或数据发布说明中明确指出这一点，以说明缺失任何原始数据的原因 [14]。

通常，研究成果传播的页面或字数限制会使所有信息难以轻松地与研究领域共享，在这种情况下，建议将这些详细信息上传到代码库或稳定的网站。对于数据集，这些内容可以简单地与数据以及模组或定制游戏源代码一起，纳入 GitHub 或其他安全的代码库中。以下内容通常不会被纳入：完整的分步操作流程、硬件和软件规格清单、完整的调查问题库，以及定制游戏或模组的源代码。为实现完全的透明度，这些内容都可以安全地上传，并在研究成果传播材料中提供链接。此外，在某些情况下，数据集本身会被赋予单独的数字对象标识符，以便于引用和访问。

6 拟议案例研究

本节详细说明一项正在进行的研究在方法设计方面的理由和考量，该研究完成后也将发布一个数据集。本案例研究通过给出真实且实用的描述，展示如何在决策过程中思考可重复性与透明度。研究人员在设计实验时所做的实际决策会因研究而异，且还有许多本文未讨论的考量；不过，本研究旨在促进整个研究过程中的开放性与透明度。

6.1 游戏和任务的选择

即将开展的实验旨在观察自然游戏过程中认知负荷的各种客观和主观测量指标之间的关系，以确定各项指标在这一情境下是否适用。本实验选择的游戏是 Minecraft，这是一款 3D开放世界程序生成沙盒游戏。Minecraft 中没有特定目标，因此本实验专门设计了一项相关任务。任务要求玩家在一名类人同伴角色的陪伴下建造房屋，该角色会提供指示、收集资源，并协助玩家建造房屋。虽然游戏原本并不包含人类同伴，但玩家与人类角色一同游戏并不显得突兀，因为 Minecraft 支持多人游戏功能，且游戏原本就包含其他类型的同伴（如猫和马）。Minecraft 因开发者提供的模组支持而广为人知，也拥有庞大的业余开发者社群，他们会制作模组并与玩家分享（包括模组的源代码）。

在本实验中，使用 Forge API，基于 Minecraft（PC 端 Java 版）1.18.1 和Forge 39.1.2对Human Companions[9] 模组进行了扩展。为该伴侣添加了许多行为，包括寻找、移除和收集木块，以及将这些木块放置到游戏环境中提供的模板上。此外，还加入了一个教程任务（在主要任务之前），要求玩家完成主要任务的伪版本，以便学习游戏操作，并熟悉最终完成主要任务所需的步骤。所有玩家指令均预先编写，并通过游戏内聊天框和语音播放提供，以确保向每位参与者提供的信息完全一致。

6.2 硬件

用于运行实验和收集数据的计算机是一台 Alien Aurora R6 台式机，搭载 Intel Core i7-7700 @ 3.60GHz 处理器、32GB RAM 和 64 位操作系统。安装了 Windows 10 Pro（版本 21H2）。显示器为 BENQ GW2760，刷新率为 60Hz，分辨率为 1920 x 1080。

[9] https://www.curseforge.com/minecraft/mc-mods/human-companions

为了收集所需的生理数据，设备在研究领域中得到良好支持，并且开发者能够为其提供长期支持，这一点非常重要。开源硬件及其配套软件采用开源模式是优先考虑的因素，因此选择了以下设备：

• **OpenBCI Ultracortex Mark IV 头戴设备**。将使用10:20电极放置系统搭建一个16 通道干式EEG系统。要实现完整的16个电极，需要主板（Cyton）和扩展板（Daisy）。购买的是中号专业组装版本。设备由一块 500 mAh 锂离子可充电电池供电，并将随附的电极替换为 5 mm 梳状可重复使用干式 Ag-AgClEEG 电极（从 OpenBCI 购买），以提高舒适度和佩戴性。数据通过 Cyton Wi-Fi 加密狗传输到计算机。

• **Emotibit**。Emotibit 使用由 400mAh 锂离子电池供电的 Adafruit Feather M0 WiFi电路板，并配有 2个 Ag/AgCL EDA 电极，用于采集多种不同的测量数据。它使用 microSD 卡记录数据，并配有用于不同佩戴位置的绑带。如图 2 所示，该设备连接在参与者的右侧肱二头肌上。• **Myoware肌肉传感器**。Myoware 肌肉传感器v1.0 测量肌电图数据，并使用 3个内置电极连接端。它连接到 Arduino Uno v3 电路板，而该电路板连接到 Adafruit 100ma 全速 USB 隔离器（连接至前述PC），以确保参与者的安全。如图 2 所示，该设备连接在参与者的左侧肱二头肌上。

• **罗技摄像头**。 一台罗技 C922 Pro Stream 网络摄像头用于录制参与者面部的视频数据。

6.3 软件

为收集上述传感器的数据及其他行为数据，研究中使用了多种软件：

• **OpenBCI GUI**。研究使用开源软件 OpenBCI GUI（v5.1.0）从 Ultracortex 收集 EEG 数据。在阻抗小组件中检查阻抗，以确保信号无噪声（对于干电极，阻抗理想情况下应低于 500 kOhms）。这些数据通过网络小组件发送至 Lab Streaming Layer。

• **Lab Streaming Layer (LSL) 和 Lab Recorder**。 为支持未来实验中的在线处理和游戏内容调整，并实现时间同步，研究使用开源程序 Lab Streaming Layer（v1.13.0），将未经预处理的数据从 GUI 传输到其配套的数据记录程序 Lab Recorder。

• **Emotibit 示波器和 LSL 标记流**。 这是 Emotibit 的开源软件，可用于流式传输和记录数据。为记录本地时间戳，示波器（v1.5.10）可以接收通过 Python 运行的 LSL 标记流，从而实现数据流的离线同步。• **摄像头视频采集**。 研究使用简单的 OpenCV（cv2 v4.7.0.68）Python 脚本启动视频流，并将第一帧的时间戳打印到视频上，以便之后同步数据流。

• **Minecraft 日志记录**。 嵌入模组代码中的是Java原生日志记录器的一个实例，用于跟踪并编码游戏中的事件。它会收集每条伴随对话（包括指令和进度消息）的时间戳（系统时钟的时间戳，精确到毫秒），以及完成整个任务所用的时间。教程任务中也会记录相同的信息（不包括进度消息）。

6.4 问卷调查

为收集参与者在任务期间感知到的认知负荷相关主观数据，任务结束后会发放两份问卷

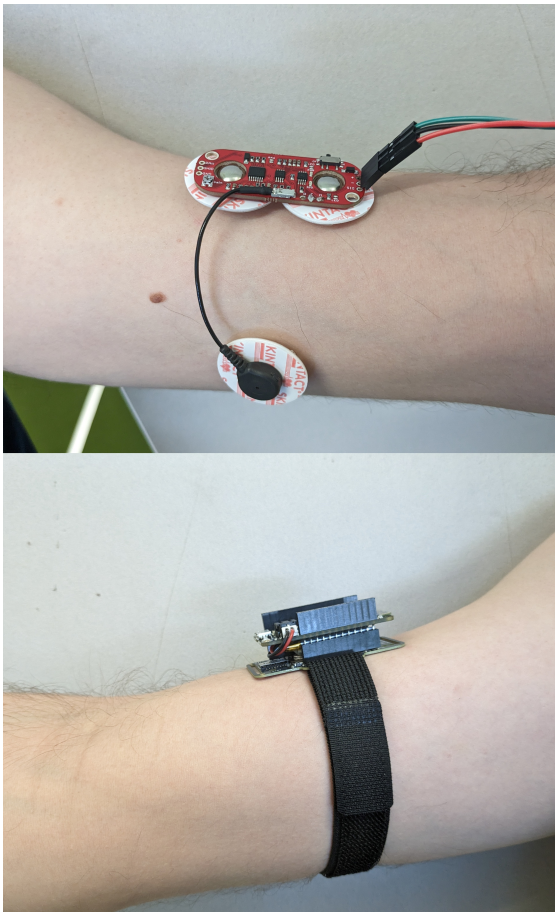


图 2：连接在参与者左侧肱二头肌上的 Myoware 肌肉传感器，以及连接在右侧肱二头肌上的 Emotibit。

参与者还被询问了他们此前使用 Minecraft 的经验。以下将说明相关理由，附录中列出了完整的问题清单：

• **电子游戏需求量表（VGDS）**。电子游戏需求量表是一种针对电子游戏的主观心理负荷测量方法，用于探究游戏需求的四个不同因素：认知、情绪、身体和社交。尽管该量表尚未被广泛使用，但本研究旨在探讨其相对于其他认知负荷测量方法的有效性，因此选用了该量表。

• **NASA-任务负荷指数（NASA-TLX）**。NASA-TLX 是一种用于评估任务负荷体验的主观问卷，包含六个单项分量表：心理需求、生理需求、时间需求、绩效、努力程度和挫败感。在过去40 年中，它被广泛应用于各种领域、语言和实验任务，并被视为测量自我报告任务负荷的黄金标准。

• **先前经验问卷**。参与者将选择最符合其 Minecraft 及 Minecraft 相关产品使用经验的回答。收集这些信息是为了控制并考虑先前经验可能对认知负荷产生的任何影响。

数据收集完成后，所有原始数据（不包括参与者的原始面部数据）将上传到GitHub代码库，感兴趣的研究人员在同意一份类似于 [24] 的使用条款表格后即可访问该代码库。数据还将

包括所有问卷材料、改编后的 Human Companions 模组，以及用于实现任何数据收集软件的代码。此永久链接将在研究成果发布时一并提供。

7 结论

本文展示了在人类用户电子游戏研究中鼓励并落实透明度，进而提升可重复性的方式。目前，面向人类用户电子游戏研究的数据集数量不多。我们希望通过提供一些切实可行的建议和推荐，促使研究人员分享其方案与数据，以便进一步探索这些现象，并就游戏过程的某些方面达成科学共识。我们建议选择模组支持完善的商业游戏，在数据收集中使用开源硬件和软件，并分享详尽的说明，以便他人复现研究设计。本文最后提出了一项正在进行中的案例研究，旨在展示如何将数据集与所有可用的研究详情及决策一并呈现。这项工作的实施不仅能提高游戏研究的整体透明度，还能通过建立科学共识，促进人类游戏研究领域更好的协作。

参考文献

[1] P. Andreoli-Versbach 和 F. Mueller-Langer。数据开放获取：理想虽受推崇，却未付诸实践。*ResearchPolicy*，43 1633，2014 9 1621( ): – 。[2] D. Banakou、P. D. Hanumanthu 和 M. Slater。白人在黑色虚拟身体中的虚拟化身会持续降低其内隐种族偏见。*Frontiers in Human Neuroscience*，第 601 页，2016 年。[3] Beat Games。Beat Saber，2019 年。[4] N. Beaudoin-Gagnon、A. Fortin-Cˆot´e、C. Chamberland、L. Lefebvre、J. Bergeron-Boucher、A. Campeau-Lecours、S. Tremblay 和 P. L. Jackson。FUNii 数据库：用于情感游戏和玩家体验研究的生理、行为、人口统计及主观电子游戏数据库。载于*2019*年第八届情感计算与智能交互国际会议 *(ACII)*，第 1–7 页。IEEE，2019 年。[5] Bethesda Softworks。上古卷轴V：天际，2011 年。[6] Bethesda Softworks。《天际》创作工具包，2012 年。[7] N. D. Bowman、J. Wasserman 和 J. Banks。电子游戏需求量表的开发。载于*Video Games*，第 208–233 页。Routledge，2018 年。[8] D. Checa 和 A. Bustillo。关于增强学习和培训的沉浸式虚拟现实严肃游戏综述。*Multimedia Tools and Applications*，79(9):5501–5527，2020 年。[9] G. Colavizza、I. Hrynaszkiewicz、I. Staden、K. Whitaker 和 B. McGillivray。将出版物与研究数据关联的引用优势。*PLOS ONE*，15(4):e0230416，2020 年。[10] C. S. Crandall 和 J. W. Sherman。论概念性重复研究在推动科学进步方面的科学优越性。*Journalof Experimental Social Psychology*，66:93–99，2016 年。[11] K. C. Elliott。科学中的透明度分类体系。*Canadian Journal of Philosophy*，52(3):342–355，2022 年。[12] Facepunch Studios Ltd. Garry’s Mod，2006 年。[13] N. I. P. S. Foundation。NeurIPS 2021 论文检查清单指南。[14] T. Gebru、J. Morgenstern、B. Vecchione、J. W. Vaughan、H. Wallach、H. D. III 和 K. Crawford。数据集数据表。*Communicationsof the ACM*，64(12):86–92，2021 年。[15] W. H. Guss、B. Houghton、N. Topin、P. Wang、C. Codel、M. Veloso 和 R. Salakhutdinov。MineRL：大型 Minecraft 演示数据集。*arXiv* 预印本 *arXiv:1907.13440*，2019 年。[16] S. G. Hart 和 L. E. Staveland。NASA-TLX（任务负荷指数）的开发：实证与理论研究结果。载于*Advances in Psychology*，第 52 卷，第 139–183 页。Elsevier，1988 年。[17] S. J¨arvel¨a、I. Ekman、J. M. Kivikangas 和 N. Ravaja。将数字游戏用作实验刺激材料的实用指南。*Transactions of the Digital Games Research Association*，1(2)，2014 年。[18] S. J¨arvel¨a, I. Ekman、J. M. Kivikangas 和 N. Ravaja。刺激物游戏。载于游戏研究方法，第 193–205 页。2015 年。[19] N. Justesen、P. Bontrager、J. Togelius 和 S. Risi。用于电子游戏游玩的深度学习。*IEEE* 游戏汇刊，12(1):1–20，2019 年。[20] K. Karpouzis、G. N. Yannakakis、N. Shaker 和 S. Asteriadis。平台游戏体验数据集。载于*2015* 年情感计算与智能交互国际会议（*ACII*），第 712–718 页。IEEE，2015 年。[21] M. Lanier、T. F. Waddell、M. Elson、D. J. Tamul、J. D. Ivory 和 A. Przybylski。虚拟现实检验：统计功效、报告结果，以及有关虚拟现实与沉浸式环境心理学研究的有效性。人类行为中的计算机，100 78，2019 70 10 1016 015 2019 06 : – 。doi: . /j.chb. . .[22]

Z. Lin、J. Gehring、V. Khalidov 和 G. Synnaeve。Stardata：星际争霸人工智能研究数据集。载于第十三届人工智能与互动数字娱乐会议，2017 年。[23] R. P. McMahan、E. D. Ragan、A. Leal、R. J. Beaton 和 D. A. Bowman。在受控实验中使用商业电子游戏的注意事项。娱乐计算，2(1):3–9，2011 年。[24] D. Melhart、A. Liapis 和 G. N. Yannakakis。唤醒度电子游戏标注（ AGAIN ）数据集。*IEEE* 情感计算汇刊，13(4):2171–2184，2022 年。[25] M. R. Mohseni、B. Liebold 和 D. Pietschmann。面向实验性游戏研究的广泛模组制作。载于游戏研究方法，第 323– 340 页。2015 年。[26] Mojang。Minecraft，2011 年。[27] J. C. Molloy。开放知识基金会：开放数据意味着更好的科学。*PLoS* 生物学，9(12):e1001195，2011 年。[28] E. 美国国家科学院、医学科学院等。科学中的可重复性与再现性。2019 年。[29] S. S. Oyelere、N. Bouali、R. Kaliisa、G. Obaido、A. A. Yunusa 和 E. R. Jimoh。探索教育类虚拟现实游戏的发展趋势：实证研究的系统综述。智慧学习环境，7(1):1–22，2020 年。[30] Roblox Corporation。Roblox，2006 年。[31] W. J. Shelstad、D. C. Smith 和 B. S. Chaparro。裂隙上的游戏体验：虚拟现实如何影响游戏用户满意度。人因与工效学学会年会论文集，61(1):2072– 2076，2017 年。doi: 10.1177/1541931213602001[32] L. Tedersoo、R. K¨ungas、E. Oras、K. K¨oster、H. Eenmaa、¨ A. Leijen、M. Pedaste、M. Raju、A. Astapova、H. Lukner 等。不同科学学科的数据共享实践和按需提供数据的可用性各不相同。科学数据，8(1):1–11，2021 年。[33] Ubisoft。刺客信条：大革命，2014 年。[34] Ubisoft。刺客信条：枭雄，2015 年。[35] B. E. Whitley Jr 和 M. E. Kite。行为科学研究原理。Routledge，2012 年。[36] C. Yildirim。虚拟现实游戏中的晕动症会降低游戏乐趣：一种中介模型。显示，59:35–43，2019 年。doi: 10.1016/j.displa.2019.07.002

表1：Minecraft体验调查问题

问题	**回答选项**
Q1以下哪项最能描述您的游戏经历? （原版）Minecraft?	“我每天都玩” “至少，我每周玩一次” “至少，我每个月会玩一次” “至少，我每年会玩一次” “我过去 5 年内至少玩过一次” “我过去 5 年里没玩过它” “从未玩过”
Q2如果你以前玩过原版 Minecraft 游戏, 你在哪些游戏主机上玩过它?	“PC” “家用游戏主机（例如 Xbox、PlayStation、Wii U）” “掌上游戏机（例如 Nintendo DS、Switch、Playstation Vita）” “移动设备（例如 Android、iPhone、iPad）” “虚拟现实（例如 Oculus、PSVR）” “其他，请注明”
Q3你玩过以下哪些 Minecraft 游戏?	“原版 Minecraft” “Minecraft：故事模式” “Minecraft Dungeons” “Minecraft Earth” “Minecraft 教育版” “其他，请注明”

表 2：NASA-TLX 问题 [16]

NASA-TLX 问题
心理需求- 这项任务在心理上有多大的要求? **(**1 = “非常**低**” 7 = “非常 **高”)**
体力需求- 这项任务的体力要求有多高? （1 = “非常**低**” 7 = “非常 **高**”）
时间压力- 任务的节奏有多急促或匆忙? （1 = “非常**低**” 7 = “非常**高**”）
表现- 你在完成被要求的任务方面做得有多成功? (1 = “**完美**” 7 =“**失败**”)
努力程度 - 为达到你的表现水平，你付出了多大努力? **(**1 = “非常**低**” 7 = “非常 **高”)**
挫败感- 你感到多么不安、气馁、恼火、紧张和烦躁? **(**1 = “非常**低**” 7 = “非常 **高”)**

表3：VGDS问题 [7]

回答选项						
1 非常不同意	2 不同意	3 有些不同意	4 既不同意也不反对	5 有些同意	6 同意	7 非常同意
认知负荷						
这款游戏对认知能力要求很高。						
玩这款游戏时，我必须非常认真地思考。						
这个游戏需要大量的脑力活动。						
这个游戏不需要付出太多脑力。						
这款游戏让我调动了全部脑力资源。						
这款游戏中的脑力挑战影响了我的游玩方式。						
这款游戏激发了我的大脑。						
情感需求						
这款游戏拨动了我的心弦。						
这款游戏触动了我的心弦。						
这款游戏打动了我。						
我与游戏内容之间有着强烈的情感联结。						
我在游戏过程中产生了许多意想不到的感受。						
体力需求						
玩完之后，我感到身体疲惫不堪。						
玩完之后，我感到身心紧绷。						
游戏过程结束后，我感到筋疲力尽。						
这款游戏对体力要求很高。						
社交需求 注：在部分陈述中提到“他人”时，可能指游戏中的其他虚拟化身、角色和/或玩家。						
社交是玩这款游戏的重要组成部分。						
玩游戏时，我会注意到游戏中的其他人。						
我不得不与游戏中的其他人互动。						
玩游戏时，我觉得自己对其他人负有责任。						
游戏中有其他人在场，会影响我的玩法。						
这款游戏对社交互动的要求很高。						