实时面部交流在社会困境中于背叛之后恢复合作

Mayada Oudah[1], * 和 John Wooders[1,2]

[1]社会科学部，纽约大学阿布扎比分校，阿联酋。[2]行为制度设计中心，纽约大学阿布扎比分校，阿联酋。

*通信请联系 mayada@nyu.edu

arXiv:2601.15211v1 [cs.GT] 21 Jan 2026

摘要

面部表情是人类交互的核心，但其在战略决策中的作用却鲜少受到关注。我们研究了实时面部交流如何影响重复社会困境中的合作。在一项实验室实验中，参与者进行重复囚徒困境博弈，实验包含两种条件：在其中一种条件下，参与者通过性别中性的虚拟形象观察对手的面部表情；在另一种条件下，则无法获得面部线索。我们利用最先进的生物识别技术实时捕捉并呈现情绪，发现面部交流显著提高了总体合作水平，尤其能促进背叛之后的合作。这种恢复性效应表明，面部表情有助于参与者以较不严苛的方式解读背叛，从而促进宽恕并恢复合作。尽管过去的行动仍然是行为最强的预测变量，但我们的研究结果凸显了面部表情在塑造战略结果方面的交流力量。这些结果为设计具有情绪响应能力的虚拟代理和数字平台提供了实践启示，使其能够在缺乏实体面对面的情况下维持合作。

引言

面部表情在人际交互中发挥着至关重要的作用。作为非语言交流的一种基本渠道，面部表情无需借助语言便能传达情绪和意图，却会显著影响他人的行为和预期 [1, 2, 3, 4, 5]。面部表情能够引导社会行为、唤起共情，并有助于调节日常及策略性场景中的合作与冲突，例如谈判、团队决策和日常交往 [6, 7, 8, 9]。在这些场景中识别和解读面部表情的能力，能够帮助个体更好地预判他人的反应，并据此调整自己的决策 [10]。

尽管非语言交流 [11, 12]发挥着核心作用，但许多关于战略决策的经济学文献都将其抽象排除在外。尤其是，经典的囚徒困境及其重复版本通常不考虑面部信号，尽管这类线索在现实世界的社会困境中 [13, 14, 15]无处不在。理解面部表情在策略互动中的作用

因此不仅具有科学意义，也具有现实重要性；在数字化世界中，面对面接触正日益被基于文本的交流所取代。

生物识别技术的最新进展使我们能够精准地研究面部表情的重要性。在本研究中，我们利用最先进的技术，通过网络摄像头，在个体进行交互时自动、准确地实时测量并分类面部表情。个体捕捉到的表情会通过虚拟形象呈现给其对手。

至关重要的是，这种方法克服了行为分析中的一个主要难题：人脸固有的信噪比问题。在标准交互中，情绪的“信号”往往会受到发送者身份这一“噪声”的干扰——种族、性别和吸引力等属性可能压过表情本身，并形成偏见 [16]。通过将生物识别数据重新映射到性别中立虚拟形象上，我们有效地将情绪信号与这些身份线索隔离开来。我们设计并评估了一组新颖的性别中立虚拟形象，以准确呈现埃克曼所定义的面部表情通用集合 [1]。

相比之下，在以往关于情绪在社会困境中作用的研究 [17, 18, 19, 20], 中，参与者会在做出决策之前或之后，从预先设定的一组面孔或状态中选择一种面部表情，或选择一段描述其情绪状态的文字。在这种情况下，参与者可能会策略性地传达自己的情绪状态。我们在本研究中通过在参与者彼此交互时实时捕捉并传递面部表情，来应对这一挑战。

在我们的实验中，参与者两人一组，进行至少30轮重复囚徒困境博弈，此后每轮继续进行的概率为90%。我们研究通过虚拟形象实时传达面部表情的能力如何影响合作率。在面部交流（FC）处理组中，玩家可以看到对手的面部表情，这些表情由一种性别中立的虚拟形象呈现，该形象反映生物识别软件检测到的最强烈情绪。在对照处理组中，即无面部交流（nFC）组，玩家不会收到视觉情绪反馈。

结果表明，观察面部表情的能力不仅有助于维持较高的相互合作率，还能显著提高背叛之后的合作率。这表明，视觉社交线索会减轻人们对非合作行为严重程度的感知，促使参与者表现出更大的宽恕。这些发现凸显了在社会困境中融入实时情绪反馈的作用：面部表情能够发挥修复作用，维持合作动态。

结果

为了评估我们设计的虚拟形象（如图1所示），我们邀请受试者参与 Qualtrics 调查问卷。在问卷中，我们请每位参与者将虚拟形象归入9个类别：8种普遍情绪（喜悦、悲伤、惊讶、恐惧、愤怒、厌恶、轻蔑、中性）以及“难以分类”类别。我们还请参与者通过拖动虚拟形象图片并将其放入他们认为合适的类别中，按照性别对虚拟形象进行分类（男性、女性、两者皆可）。图2展示了参与者的分类准确率。除轻蔑 （准确率为 $\approx$ 57%）外，所有虚拟形象的分类准确率均超过80%。这一较低的准确率与既有文献一致；文献指出，轻蔑常与厌恶或愤怒混淆，且难以识别 [21]。事实上，该虚拟形象的大多数误分类都落在“愤怒”“厌恶”和“难以分类”类别中（见补充信息图S1），这证实了尽管存在语义歧义，该虚拟形象仍成功传达了预期的负效价。图3展示了参与者的性别分类。除轻蔑虚拟形象外，对于每个虚拟形象，大多数参与者都将其性别归为“两者皆可”。

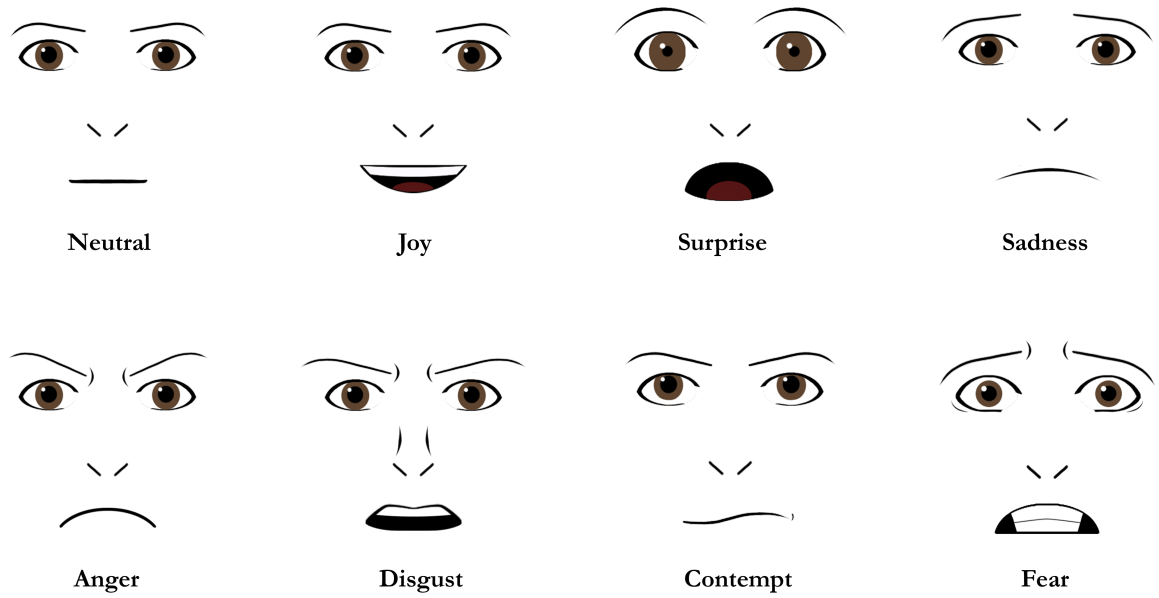


图1：**用于表现面部表情的虚拟形象**。这些虚拟形象采用示意性设计，以最大限度地减少“恐怖谷”效应，同时最大限度地提高清晰度。

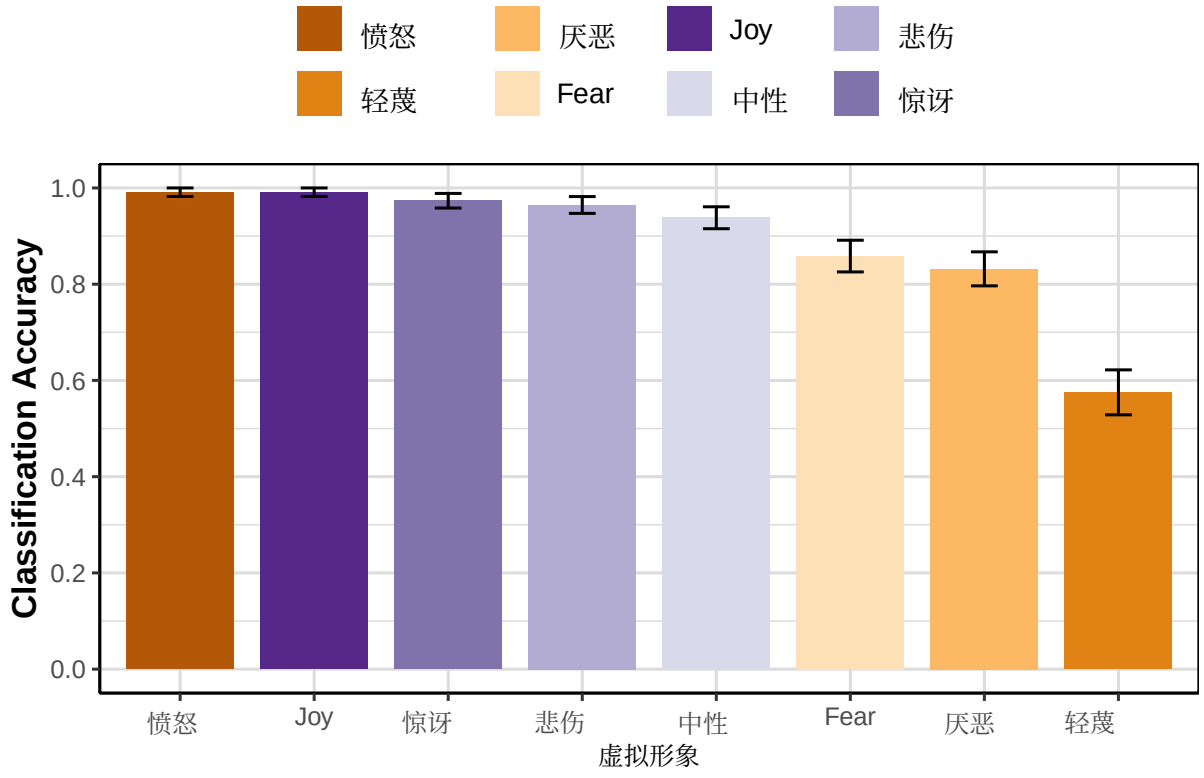


图2：**虚拟形象分类准确率**。误差线表示均值标准误。

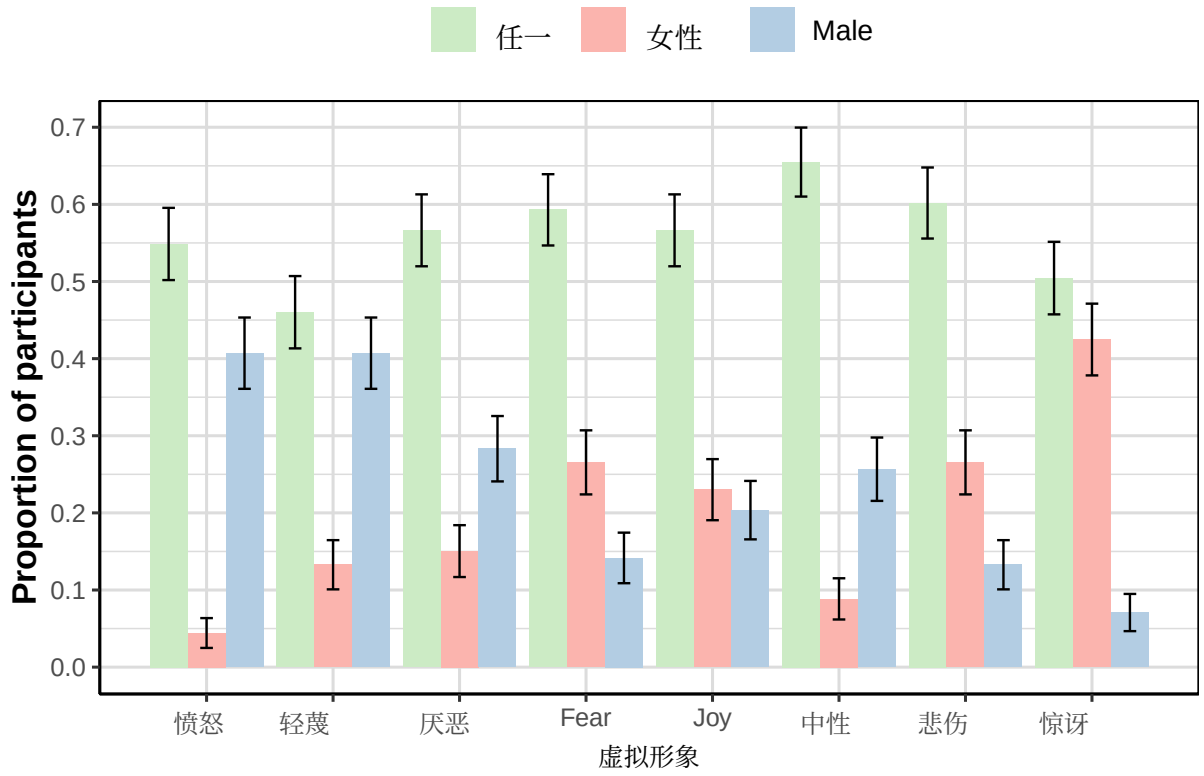


图3：**虚拟形象的性别分类**。误差线表示均值标准误。

鉴于虚拟形象的分类准确率很高（轻蔑除外），且最常见的性别分类是“两者皆可”，我们使用这些虚拟形象来呈现参与者在囚徒困境实验中的面部表情。该实验包含两种处理条件，区别在于玩家是否能看到对手的虚拟形象。我们将启用面部交流的处理条件称为FC，将禁用面部交流的处理条件称为nFC（详情请参见方法部分）。在FC处理条件下，面部表情通过虚拟形象传达。每位玩家都能在自己的屏幕上实时看到另一位玩家的面部表情。

图4展示了两种处理条件下的合作率。与nFC处理条件相比，参与者在FC处理条件下的合作程度更高。通过回归分析验证了这一结果，并采用广义线性混合模型（GLMM）来处理被试内依赖关系和重复测量。图5展示了回归系数（完整模型详情见补充表S1）。FC的正系数（$\beta = 0.47$, $p < 0.001$）表明，与匿名基线相比，面部虚拟形象的存在显著提高了合作率。

为了解感知到的面部表情对合作决策的影响，我们以参与者是否合作为因变量进行回归分析，考察：（i）作出决策前自己和另一位玩家所显示的面部表情；以及（ii）上一轮双方玩家的合作决策。图6显示，虽然喜悦（自己和另一位玩家）的出现是合作的正向预测变量，但近期行为历史（自己和对手先前是否合作）是当前行为最强的预测变量，这证实了以牙还牙等互惠策略的主导作用。完整模型详情见补充表S2。

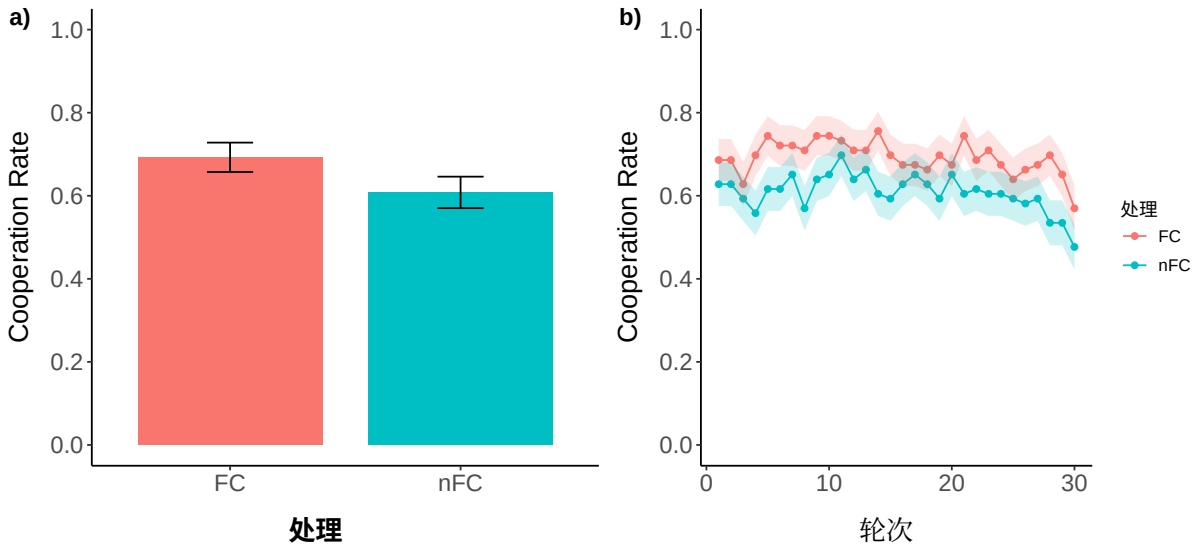


图4：**实验结果——合作。a)** 展示各处理条件下的合作率。误差线表示均值标准误。**b)** 展示了30轮中的合作率。

我们进一步分析参与者的面部表情是否会受到自己和对手行动的影响。关键的是，该分析依据iMotions软件捕捉到的原始生物特征值，这些数值不同于显示给对手的面部表情。与应用60%强度阈值以确保信号清晰度的虚拟形象不同，这些测量值能够捕捉细微的情绪变化和内在反应，而这些可能并未被另一位玩家看到。

图7展示了参与者自身的行动如何影响这些内部情绪状态。显然，与nFC处理相比，参与者在FC处理中的情绪参与度更高（表达了更多非中性情绪）。参与者合作时，往往会表达积极情绪，尤其是喜悦，这表明他们感到情绪上的满足，或是在传递亲社会信号（见图7 (a)和(c)）。这与合作能带来情绪奖励的观点一致。

在FC处理中，自身背叛与较少的喜悦以及消极情绪的增加相关（见图7 (b)）。这种生物特征值骤升可能表明存在内心冲突、内疚或社会不满——即使表情过于细微，未能触发虚拟形象，这些情感反应仍会发生。相比之下，在nFC处理中，自身背叛与背叛者测得的情绪状态没有任何显著变化相关，这凸显了匿名互动中较低的情绪利害程度。

图8展示了对手的行动对参与者测得的内部情绪状态的影响。与自身行动的研究结果一致，FC处理中的参与者表现出的情绪反应性显著高于nFC处理中的参与者。

当对手合作时，参与者的喜悦情绪会明显上升，反映出社会认同、感激之情或共同的亲社会取向（见图8（a）和（c））。相反，当对手背叛时，FC处理中会触发强烈的负面生物特征信号，具体表现为轻蔑和悲伤情绪骤升（见图8（b））。这种强烈的负面反应（在匿名条件下并未出现）表明，当面部身份信息可见时，背叛不仅会被视为策略性损失，也会被视为一种社会或道德违规。

至关重要的是，情绪状态时间线揭示了预期方面的差异。在nFC处理中，对手最终背叛的回合中，“事前喜悦”（即结果揭晓前测得的喜悦程度）显著较低（见图8（d））。这表明，在匿名博弈中，背叛往往是现状；参与者被困在相互背叛和负面情感的循环中，并准确预料到对方会再次背叛。

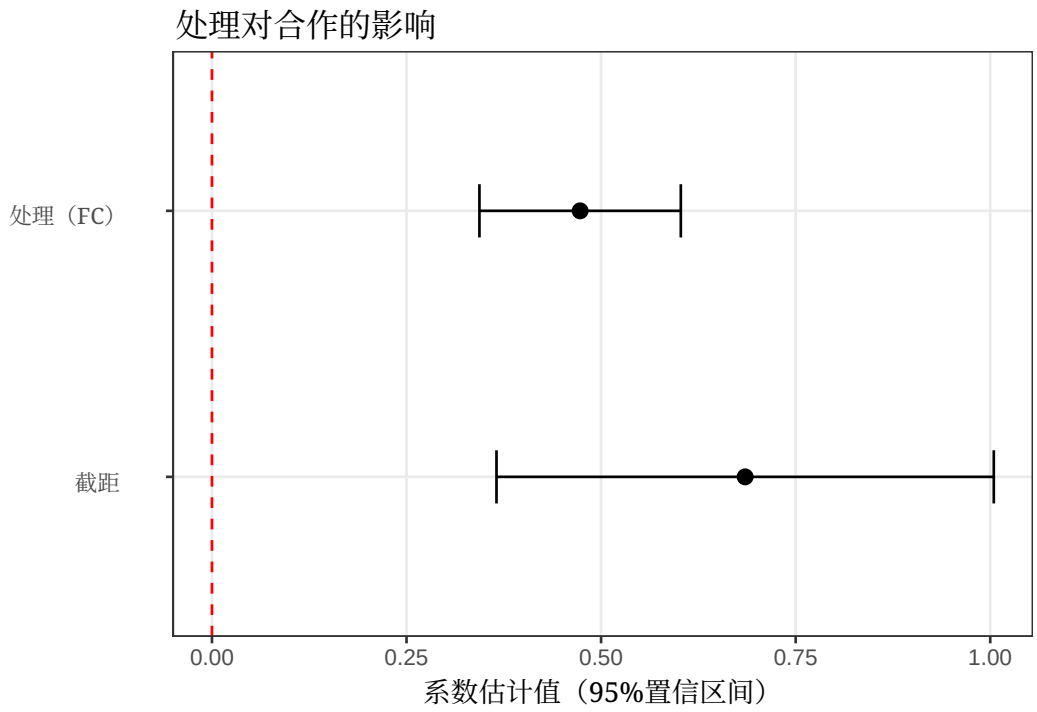


图5：**面部交流显著提高合作的可能性**。预测合作的广义线性混合模型（GLMM）固定效应系数（$N = 5,\ 160$）。点估计表示相对于非面部交流基线的对数优势；误差线表示95%置信区间。

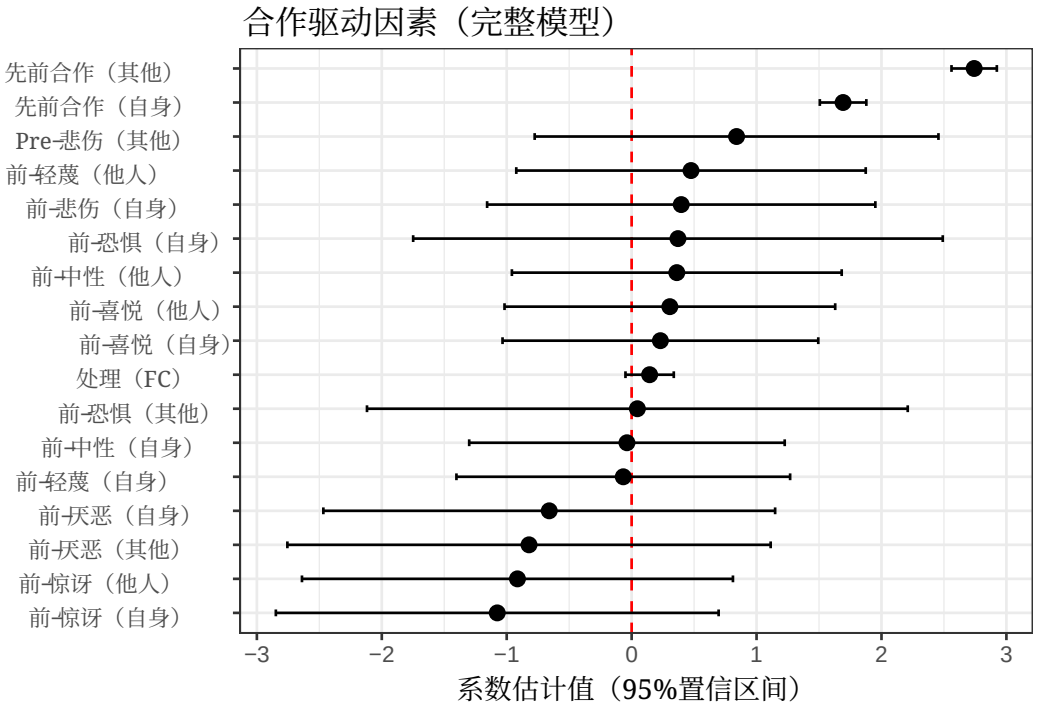


图6：**游戏历史是合作最强的预测变量**。 广义线性混合模型对结构变量和情绪变量的估计。点表示对数优势；误差线表示95%置信区间。

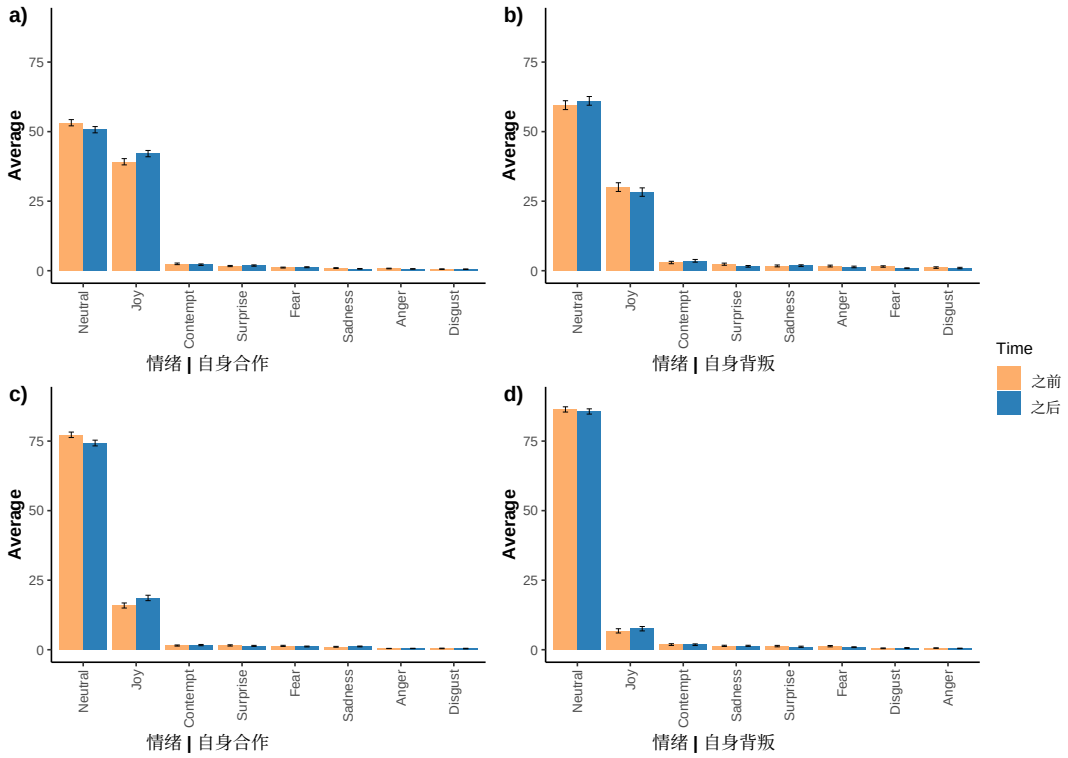


图7：**自身行动前后的情绪测量值**。误差线表示均值标准误。

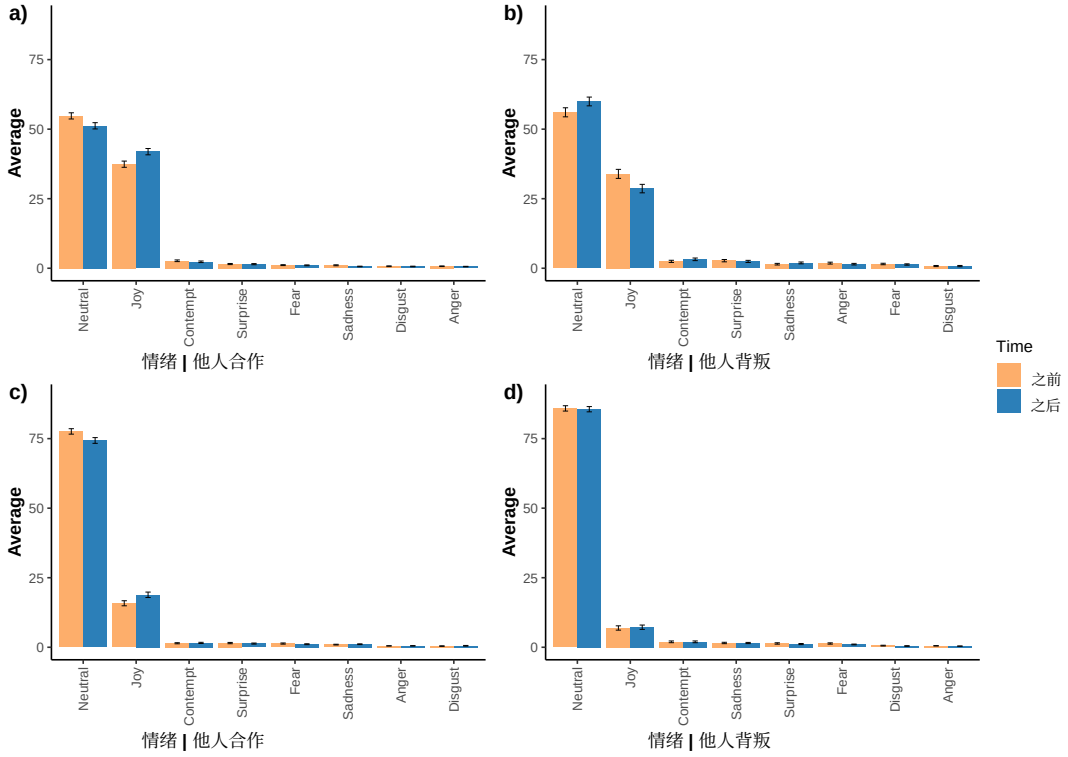


图8：**另一位玩家行动前后的情绪测量值**。误差线表示均值标准误。

相比之下，在FC处理条件下，即使在对手随后背叛的回合，事前喜悦仍然很高。这表明，直到背叛发生的那一刻，FC组参与者始终保持着情绪乐观或合作意图。因此，FC处理条件下的背叛违背了这种持续的希望，也解释了随后强烈的修复努力（宽恕）。更多信息请参阅我们的补充信息。

图9（a）显示，在上一回合采取合作后，参与者在FC和nFC处理条件下的合作率。同样，图10（a）显示，在对手上一回合采取合作后，参与者在两种处理条件下的合作率。在这两种情况下，处理条件之间的合作率均无统计显著差异。

不过，如果玩家在上一回合背叛，两种处理条件之间就会出现显著差异。图9（b）显示，在上一回合自身背叛后，参与者在FC处理条件下比在nFC处理条件下更有可能合作。玩家之间进行面部交流时，参与者在自身背叛后可能会感到内疚和悔恨；没有面部交流时则不会有这种感受。图10（b）显示，在对手背叛后，参与者在FC处理条件下也显著比在nFC处理条件下更有可能合作。在存在面部交流的情况下，对手背叛后，参与者表现得更加宽容，也更愿意合作。这表明，视觉社交线索（面部表情）在维持合作动态方面发挥着修复作用。

实验结束时，每位参与者都完成了一份调查问卷，对另一位玩家的可喜欢程度、智力、合作性、可信度、自私程度、宽恕程度、欺凌倾向，以及是否愿意再次与其互动进行评分（更多详情见图11（a）和补充材料）。我们发现，两种处理条件下，参与者对另一位玩家“宽恕程度”的评价存在统计显著差异。与nFC组参与者相比，FC组参与者认为另一位玩家更宽容（见图11（b））。这一结果支持面部交流能在背叛之后促进宽恕——见图10（b）。

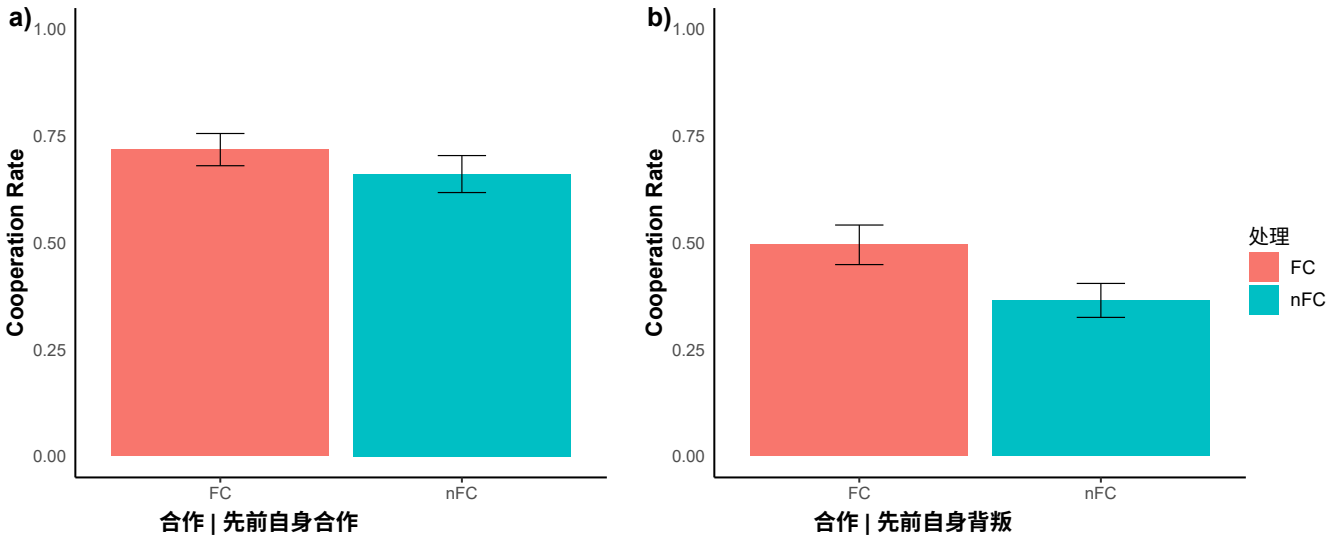


图9：**基于自身先前决策的条件合作。a）**上一回合自身合作时的合作率，以及**b）**上一回合自身背叛时的合作率。误差线表示均值标准误。

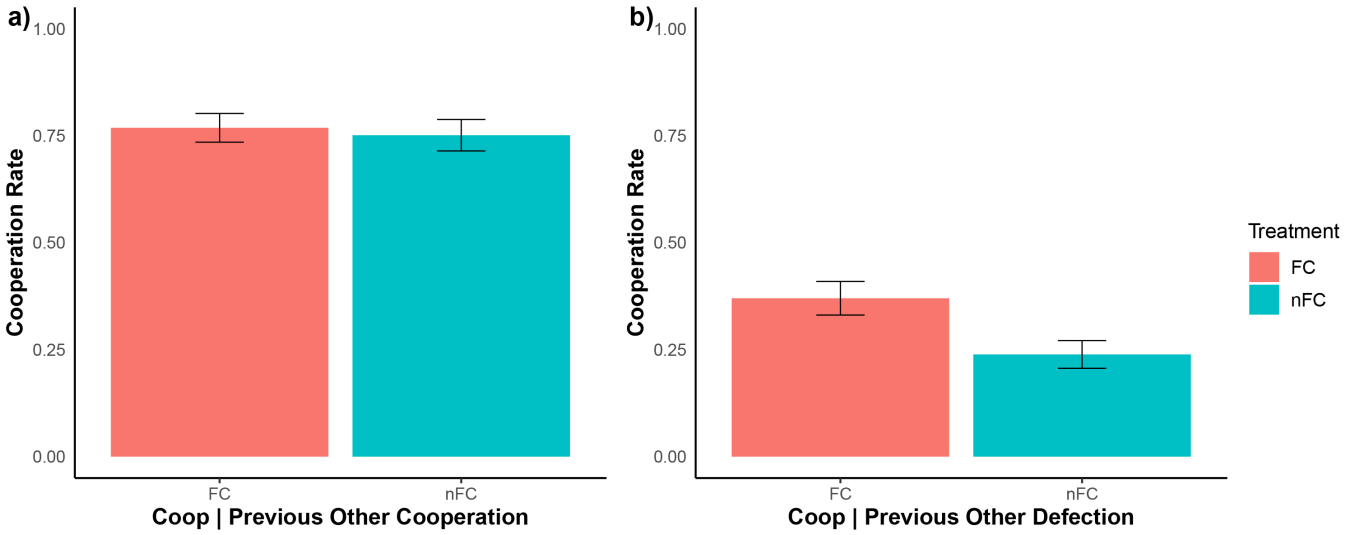


图10：**另一名玩家先前决策下的条件合作。a**）另一位玩家在上一回合合作时的合作率，以及**b**）另一位玩家在上一回合背叛时的合作率。误差线表示均值标准误。

讨论

我们的结果表明，通过虚拟形象进行面对面交流能够促进社会困境中的合作，并从根本上改变人们对背叛的看法。在nFC处理中，背叛不过是一个冷冰冰的数字选择，标准的以牙还牙逻辑（“你背叛，我也背叛”）占据主导地位。但如果参与者能在对手背叛之前和之后看到对手的虚拟形象，他们就能获得额外的社会与情感信息，从而更宽容地应对对手的背叛。也就是说，在FC处理中，人们可能会注意到某种表情——例如，背叛之后的悲伤，或合作之后的喜悦。这些面部线索表明，背叛可能并非出于故意，或者背叛者对此感到愧疚，从而使参与者更愿意再给对方一次机会。相比之下，在nFC处理中，参与者只能看到结果，无法看到任何虚拟形象。没有代价高昂的道歉信号来缓和冲击，因此他们比FC处理中更倾向于报复。

无论是自己还是对手背叛，面部交流的影响在背叛之后尤为显著。在标准的重复博弈中，背叛通常会引发恶性循环，互惠机制会惩罚背叛。然而，在这里，面部表情似乎缓和了这种反应的僵化。FC处理中的玩家在合作遭到破坏后更有可能重新建立合作，这表明面部线索为行动提供了解读背景，否则这些行动可能会被视为纯粹出于自利且充满敌意。

面部表情发挥着情感上的“修复工具”作用，帮助参与者将悔意或其他动机归因于背叛，而这种背叛原本可能会被视为蓄意行为。这与相关研究的发现一致：情绪——尤其是通过面部表情传达的情绪——不仅具有反映作用，也具有沟通作用；它们传达意图、引发共情并促进和解 [4, 5, 19]。

尽管游戏历史确立了合作基线，但我们的分析揭示，面部表情仅在这一基线遭到破坏时，才会发挥关键的“断路器”作用。虽然过去的行动仍是常规行为的主要驱动力——这与标准互惠原则一致——但在发生违约后，情感线索会变得至关重要。通过重新诠释意外背叛，面部表情

调节了僵化的以牙还牙逻辑，使玩家能够优先选择宽恕而非报复。因此，记忆维持稳定性，而情感沟通则恢复稳定。

重要的是，FC处理中的玩家不仅表现得更合作，情绪投入也更高。他们的面部表情，尤其是喜悦和轻蔑，更为明显，也更能对自身及他人的行动做出反应。表现力的提升表明，被他人看见以及看见他人，会增强交互的社会显著性并促进合作。

这些发现具有若干实践意义。在通常缺乏面部反馈的数字环境中，基础的虚拟形象驱动的情绪表达可以重新唤起问责、共情和共同理解等要素。对于虚拟代理和人工智能系统的设计者与开发者而言，我们的结果凸显了嵌入实时情绪反馈回路的重要性：这些回路应当可解释且能及时响应，而不只是被动反应。

我们的实验设计涉及人类情感数字中介所固有的特定权衡。虽然实时情绪映射确保了严格的实验控制，但设定 60% 的强度阈值可能会过滤掉细微的微表情。通过抑制低强度信号（例如低于阈值的淡淡微笑），我们的设计优先考虑信号清晰度，而非敏感度，以防止虚拟形象闪烁或信号不稳定。因此，我们的结果可能保守估计了面部表情的修复力量，因为细微的和解线索实际上遭到了屏蔽。尽管虚拟形象具有人工属性，但图11（a）中较高的可喜欢程度评分表明，参与者将这种交互视为真实的社会交流，而非单纯的人机交互任务。

总而言之，我们的研究结果表明，在受控条件下，即使是细微的面部交流，也能显著改变策略环境中的行为结果。面部表情改变了人们对背叛的解读方式，让我们得以窥见意图，并为情绪修复提供了途径。未来的研究可以探讨这些机制在不同文化背景、群体规模或高风险环境中如何发挥作用，也可以进一步考察旨在表达

类似人类的情绪同样能够促进信任与合作。在日益数字化的世界中，理解情绪可见性如何塑造长期动态尤为及时——“看见他人并被他人看见”的能力，可能不仅决定我们的选择，还决定我们在人际关系中成为怎样的人。

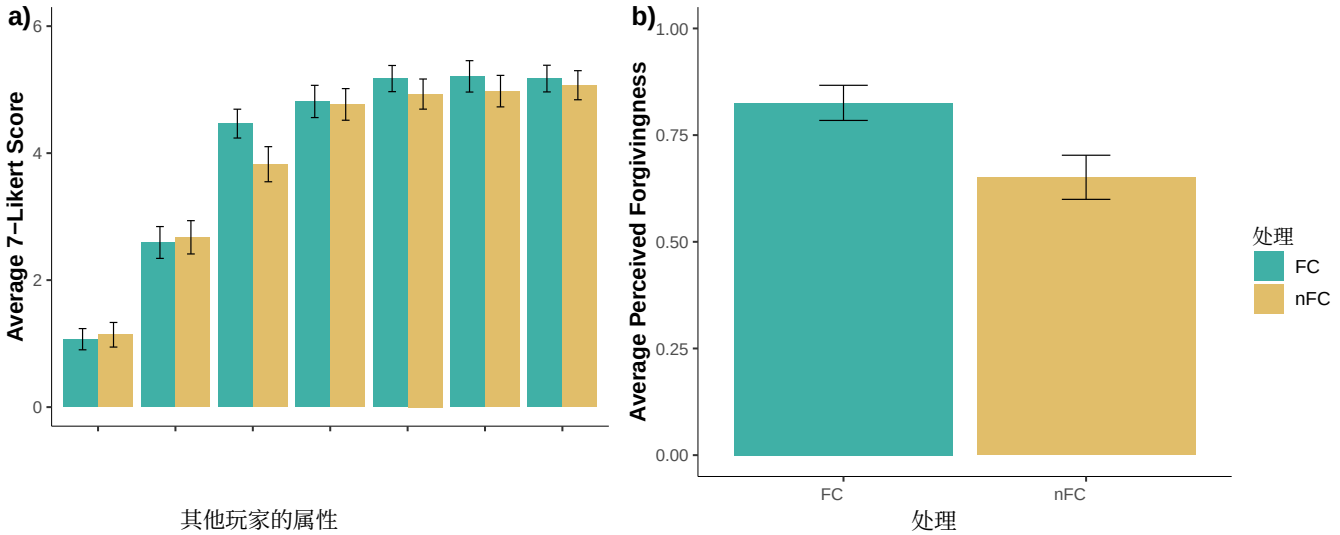


图11：**参与者对另一位玩家属性的感知。a）**另一位玩家各项属性的平均感知得分。**b）**参与者报告另一位玩家的宽恕程度达到李克特量表得分3分或以上的频率。误差线表示均值标准误。

方法

我们的所有实验方案均获得了纽约大学阿布扎比分校（NYUAD）机构审查委员会的批准。在第一个实验中，受试者评估了设计的化身的准确性。第二个关于互动式囚徒困境的实验采用被试间设计，参与者在不同日期接受不同的处理条件招募。参与者并不知道存在不同的处理条件。受试者在参与前签署了知情同意书。

实验设计与设置

为了探究面部表情对决策的影响，我们首先设计了一组自己的中性别化虚拟形象，用来呈现埃克曼 [1] 所确定的 8 种普遍情绪：中性、悲伤、快乐、恐惧、愤怒、惊讶、厌恶和轻蔑。[1] 图1展示了所创建的虚拟形象。我们刻意采用示意图风格的虚拟形象，以避免“恐怖谷”效应，即过于逼真但不够完美的虚拟形象会引起反感 [22, 23]。展示虚拟形象而非玩家的真实面孔，可以消除面部特征（如性别、族裔、吸引力等）可能带来的决策偏差[24, 25, 26]。为验证这些设计的化身是否准确，在第一个实验中，受试者通过将虚拟形象归类为相应的普遍情绪，并判断其性别（男性、女性、两者皆可），对这些虚拟形象进行评估。

我们选择研究囚徒困境，因为这是一种引人入胜的游戏，玩家可能会展现丰富多样的面部表情。在实验中，我们使用了 Blake 等人 [27] 的图形用户界面（GUI）（见图12）。游戏中，两名玩家相对而坐，参与者是位于下方的玩家，对手则是位于上方的玩家。每位玩家都可以点击相应按钮来“推”或“拉”自己的托盘。游戏以同步方式进行，因为玩家在做出自己的行动之前无法观察对手的行动。游戏由固定配对进行。

我们在原有图形用户界面的基础上进行了扩展，在屏幕中央添加了一个区块，通过虚拟形象呈现对手的面部表情。具体而言，只要生物识别软件测得的最强情绪强度超过60%的阈值，我们就显示与该情绪相对应的虚拟形象；否则显示中性虚拟形象。如果软件未检测到面部，该区块则为空。

实验包含两种处理条件：(i) 面部交流（FC）和 (ii) 无面部交流（nFC）。在两种处理条件下，参与者的面部表情均通过 iMotions [28], 进行实时捕捉和测量；该软件使用 Affectiva Affdex [29, 30], 能够准确、自动地进行面部表情分析。该软件通过安装在每台显示器顶部的网络摄像头检测并测量面部表情。在 FC处理中，参与者在整个实验过程中都能看到其对手的虚拟形象。在 nFC处理中，参与者的面部表情会被记录，但不会显示。有关 iMotions 平台的更多详情，请参阅补充材料。

实验在纽约大学阿布扎比分校（NYUAD）的社会科学实验室（SSEL）举行，参与者分别坐在独立的隔间里。我们每场邀请 8 名参与者参加。参与者阅读并签署知情同意书后，实验员讲解了实验说明，随后参与者在电脑上完成教程，以熟悉游戏中的行动操作方式，并根据自己的面部表情校准软件。参与者获知，在囚徒困境博弈的整个过程中，他们的面部表情都会被记录和分析，并被要求不要遮挡面部。参与者还获知，他们将与同一位固定参与者至少进行 30 轮囚徒困境博弈，此后每一轮结束时，交互都有 10% 的概率终止

[1] 由于现有文献中的虚拟形象往往带有性别特征，因此我们设计了自己的虚拟形象。

。在这种设定下，采用冷酷策略或以牙还牙等策略，可以使每轮合作成为子博弈完美均衡。参与者还获知，实验结束时，他们将获得随机抽取的两轮的报酬。

呈现说明后，实验员逐一照看每位参与者，确保他们坐姿合适，以便网络摄像头（Logitech C920 Widescreen HD Pro）准确捕捉其面部；同时请他们做出几种面部表情，以确认 iMotions 模块能够检测到这些面部表情。iMotions 完成校准后，参与者被要求在游戏软件中输入玩家编号（实验员在他们进入实验室时提供）、年龄和性别。配对玩家输入信息后，游戏便开始。游戏进行期间，iMotions 软件在后台运行。参与者每轮做出决定没有时间限制。游戏持续约 10–15 分钟，具体时长取决于实际进行的回合数以及参与者做决定的速度。交互结束后，屏幕上会显示总收益。参与者填写了一份实验后调查问卷，内容涉及他们在游戏中如何做出决定、人口统计信息，以及他们如何评价另一位玩家的行为。参与者通过 YouGotaGift 获得报酬。YouGotaGift 是一家中东数字礼品卡公司，支持 700 多个品牌的电子礼品卡，其中包括亚马逊。



图12：**用户界面和处理条件**。**a**）FC处理下双人囚徒困境游戏的界面，玩家可以观察另一位玩家的虚拟形象。参与者控制位于底部的玩家，并看到位于对面的另一位玩家。每位玩家都可以点击相应的行动按钮，推动（合作）或拉动（背叛）杠杆。推动杠杆会给另一位玩家三枚硬币，并使一枚硬币掉入黑洞。拉动杠杆会给该玩家一枚硬币，并使三枚硬币掉入黑洞。**b**）nFC处理下的界面。规则与FC处理相同，但玩家无法观察另一位玩家的虚拟形象。**c**）相互合作的界面。**d**)合作对抗背叛者的界面。

招募

我们通过 ORSEE 招募系统招募了纽约大学阿布扎比分校年满 18 岁的学生，请他们评估设计的化身并参与我们的实验研究。要符合参与囚徒困境博弈的资格，学生不能留有浓密胡须或佩戴眼镜，以避免影响 iMotions 收集的数据质量。评估过设计的化身的学生，只要同时满足这两项资格标准，也可以参与囚徒困境博弈。同一名学生不得同时参与 FC处理和 nFC处理。

调查问卷的数据收集时间为2022年10月 14^{th} 日至11月 3^{rd} 日。FC处理的数据收集时间为2023年5月 29^{th} 日至6月 16^{th} 日，nFC处理的数据收集时间为2023年11月 6^{th} 日至 24^{th} 日。我们为虚拟形象评估调查招募了113名参与者，并为囚徒困境博弈共招募了172名参与者，其中FC处理和nFC处理各86名。

参与者完成评估调查后获得15阿联酋迪拉姆（4.08美元），平均完成时间为8.60分钟。囚徒困境博弈的参与者除获得30阿联酋迪拉姆（8.17美元）的出席费外，还可获得0阿联酋迪拉姆（0美元）至190阿联酋迪拉姆（51.73美元）不等的奖金，具体金额取决于他们的决策以及随机抽选用于支付的轮次。FC处理的平均报酬为121.86阿联酋迪拉姆（33.18美元），nFC处理的平均报酬为116.02阿联酋迪拉姆（31.59美元）。平均完成时间为20分钟。

受试者特征

完成虚拟形象评估调查的参与者中，女性占56.64%，男性占42.48%；其中36.28%来自南亚（16.81%来自欧洲和中亚，14.16%来自东亚和太平洋地区，11.50%来自中东和北非，10.62%来自撒哈拉以南非洲，6.19%来自拉丁美洲和加勒比地区，3.54%来自北美洲），平均年龄为20.97岁（标准差 = 1.38）。完成FC实验处理的参与者中，男性占60.47%，女性占39.53%，平均年龄为20.31岁（标准差 = 1.51）。完成nFC实验处理的参与者中，男性占52.33%，女性占46.51%，其他占1.16%。参与者的平均年龄为20.27岁（标准差 = 1.46）。

数据可用性

这些研究中分析的所有数据将在论文发表时存入开放科学框架，并可应要求提供审阅。

代码可用性

绘图、回归分析和统计分析均使用 R（版本：4.3.2）执行。代码将在论文发表时存放于开放科学框架，并可应要求提供审阅。

参考文献

[1] Ekman, P. 基本情绪论证。《认知与情绪》**6**，169–200（1992）。[2]Frith, C. 面部表情在社会互动中的作用。《皇家学会哲学汇刊：生物科学》**364**，3453–3458（2009）。[3] Knapp, M. L.、Hall, J. A. & Horgan, T. G.《人际互动中的非言语沟通》（第 8 版）（马萨诸塞州波士顿：圣智学习，2013）。[4] Sagliano, L.、Ponari, M.、Conson, M. & Trojano, L. 编者按：情绪的人际影响：面部表情对社会互动的影响。《心理学前沿》**13**（2022）。[5] van Kleef, G. A. & Cˆot´e, S. 情绪的社会效应。《心理学年度评论》**73**，629–658（2022）。[6]Hatfield, E.、Cacioppo, J. T. & Rapson, R. L. 情绪传染。《心理科学最新进展》**2**，96–99（1993）。[7] Holland, A. C.、O'Connell, G. & Dziobek, I. 面部模仿、共情与情绪识别：相关性的荟萃分析。《认知与情绪》**35**，150–168（2021）。[8] Kleef, G. A. V. 情绪如何调节社会生活：情绪即社会信息（EASI）模型。《心理科学最新进展》**18**，184–188（2009）。[9] Centorrino, S.、Djemai, E.、Hopfensitz, A.、Milinski, M. & Seabright, P. 信任互动中的诚实信号：被评定为真实的微笑会诱发信任，并传递更高获利机会的信号。《进化与人类行为》**36**，8–16（2015）。[10] Ekman, P.《情绪的真相：识别人脸与情感，以改善沟通和情绪

al life

（时代图书公司／亨利·霍尔特出版公司，2003）。[11] Sally, D. 同情、读心与社会交互的一般理论及其在囚徒困境中的应用。《社会科学信息》**39**，567–634（2000）。[12] Camerer, C. F.《行为博弈论：策略互动实验》（普林斯顿大学出版社，新泽西州普林斯顿，2003）。[13] Axelrod, R.《合作的进化》（基础图书，纽约，1984）。[14]Fehr, E. & G¨achter. 公平与报复：互惠经济学。《经济展望杂志》**14**，159–181（2000）。[15] Binmore, K.《真实博弈：博弈论文本》（牛津大学出版社，牛津，2007）。[16] Stirrat, M. & Perrett, D. 合作与信任的有效面部线索：男性面部宽度与可信度。《心理科学》**21**，349–354（2010）。[17] Hopfensitz, A. & Reuben, E. 情绪对社会惩罚有效性的重要性。《经济学杂志》**119**，1534–1559（2009）。

[18] Mussel, P., G¨oritz, A. S. & Hewig, J. 微笑的价值：面部表情会影响最后通牒博弈中的反应。《判断与决策》**8**，381–385（2013）。[19] de Melo, C. M., Gratch, J. & Carnevale, P. J. 人类与计算机：情绪表达对人类决策的影响。《IEEE情感计算汇刊》**6**，127–136（2015）。[20] de Melo, C. M. & Terada, K. 情绪表达与策略在促进重复囚徒困境中的合作方面的相互作用。《科学报告》**10**（2020）。[21] Matsumoto, D. & 埃克曼. 表情、标签与轻蔑描述之间的关系。《人格与社会心理学杂志》**87**，529–40（2004）。[22]Mori, M., MacDorman, K. F. & Kageki, N. 恐怖谷 [来自现场]。《IEEE机器人与自动化杂志》**19**，98–100（2012）。[23] Oudah, M., Makovi, K., Gray, K., Battu, B. & Rahan, T. 经验感知影响利他行为，能动性感知影响人机交互中的信任。《科学报告》**14**（2024）。[24] Wilson, R. K. & Eckel, C. C. 以貌取人：信任博弈中的美貌与预期。《政治研究季刊》**59**，189–202（2006）。[25] Stirrat, M. & Perrett, D. 合作与信任的有效面部线索：男性面部宽度与可信度。《心理科学》**21**，349–354（2010）。[26] Chen, J. et al. 加工面部吸引力的电生理相关指标及其对合作行为的影响。《神经科学快报》**517**，65–70（2012）。[27] Blake, P. R., Rand, D. G., Tingley, D. & Warneken, F. 未来阴影促进儿童在重复囚徒困境中的合作。《科学报告》**5**（2015）。[28] iMotions A/S. iMotions软件，9.0版（2026）。网址：https://imotions.com/。[29] St¨ockli, S., Schulte-Mecklenbeck, M., Borer, S. & Samson, A. C. 使用Affdex和FACET进行面部表情分析：一项验证研究。《行为研究方法》**50**，1446–1460（2018）。[30] Kulke, L., Feyerabend, D. & Schacht, A. 比较affectiva imotions面部表情分析软件与肌电图在识别情绪面部表情方面的效果。《心理学前沿》**11**（2020）。

致谢

作者衷心感谢 Rebecca Morton 教授（1954–2020）在本研究构思初期所作出的奠基性贡献。尽管她在本研究最终实施之前便已离世，但她的思想与指导对研究早期方向的确立起到了关键作用。

本研究得到 ASPIRE 卓越研究奖（AARE-2019-48）的资助。Wooders 衷心感谢 Tamkeen 通过纽约大学阿布扎比分校研究院奖 CG005 提供的资金支持。

作者贡献声明

M.O. 构思了这项研究。M.O. 和 J.W. 设计了实验。M.O. 开展了实验并收集了数据。M.O. 和 J.W. 分析了数据、解读了结果，并撰写了手稿。

利益冲突声明

作者声明不存在利益冲突。

补充信息 - 实时面部交流在社会困境中背叛之后恢复合作

Mayada Oudah[1,*] 和 JohnWooders[1,2]

[1]纽约大学阿布扎比分校社会科学部，阿联酋。[2]纽约大学阿布扎比分校行为制度设计中心，阿联酋。

*通信请联系 mayada@nyu.edu

目录

1. 补充方法 2. 补充图
3. 补充表

1 补充方法

实验设计与收益

参与者进行了一种互动式重复囚徒困境实验，实验具有无限期时域（在 30 轮之后随机以 10% 的概率终止）。图 S1 至 S5 展示了该游戏的图形用户界面及所有可能的结果。在每一轮中，每位参与者都可以选择合作（推动杠杆）或背叛（拉动杠杆）。收益（以实验积分计）由标准收益矩阵结构决定： $T > R > P > S$ 和 $2R > T + S$。
实验中使用的具体参数如下：

- 奖励（R）： 60 分（相互合作）
- 诱惑（T）： 80 分（背叛合作者）
- 惩罚（P）： 20 分（相互背叛）
- 受骗者（S）： 0 分（合作对抗背叛者）

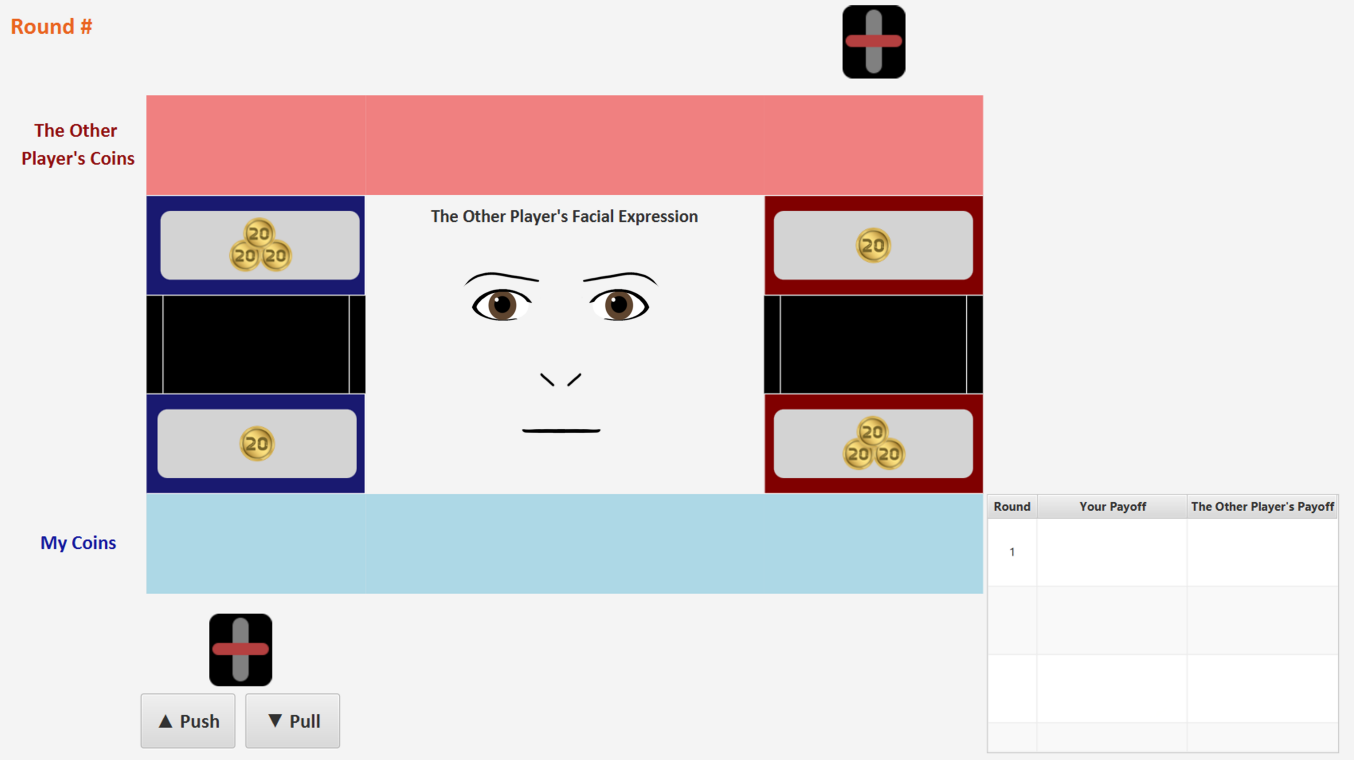


图 S1：**FC 条件下的游戏图形用户界面**。参与者看到自己是位于左下方的玩家，而另一位玩家位于对面右端。每位玩家都可以通过点击相应的行动按钮，使用杠杆推动（合作）或拉动（背叛）自己的托盘。推动托盘会向另一位玩家交付三枚硬币，并使一枚硬币掉入黑洞。拉动托盘会向玩家交付一枚硬币，并使三枚硬币掉入黑洞。nFC 条件下，中间区块为空白。

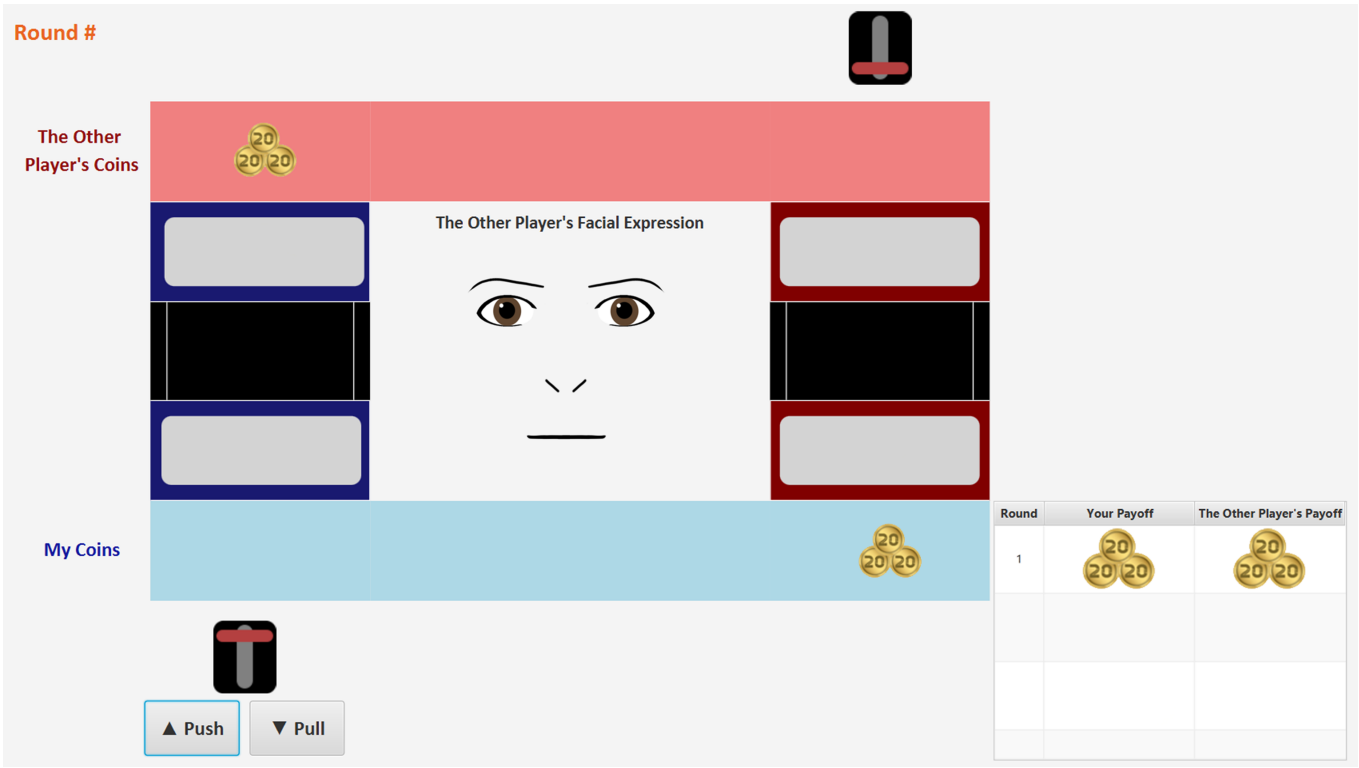


图 S2：图形用户界面：相互合作。

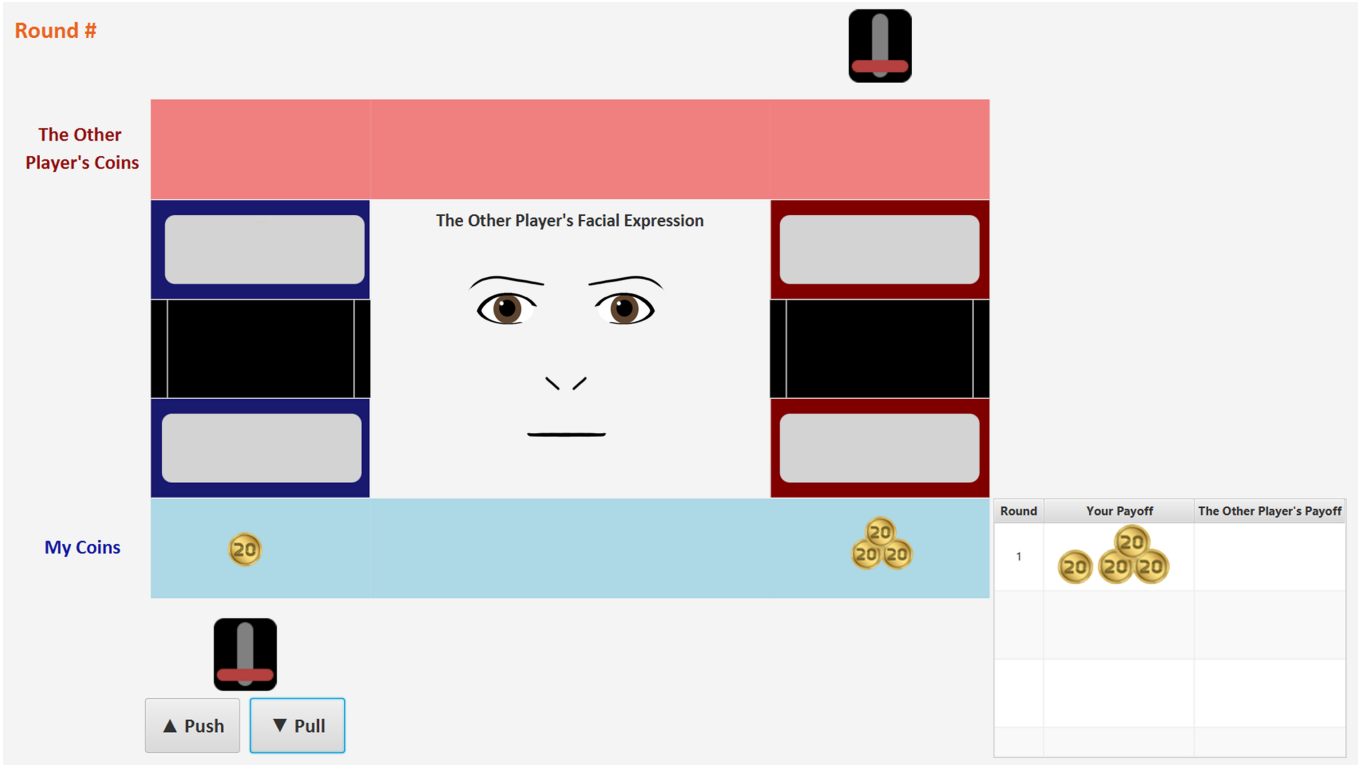


图 S3：图形用户界面：背叛合作者。

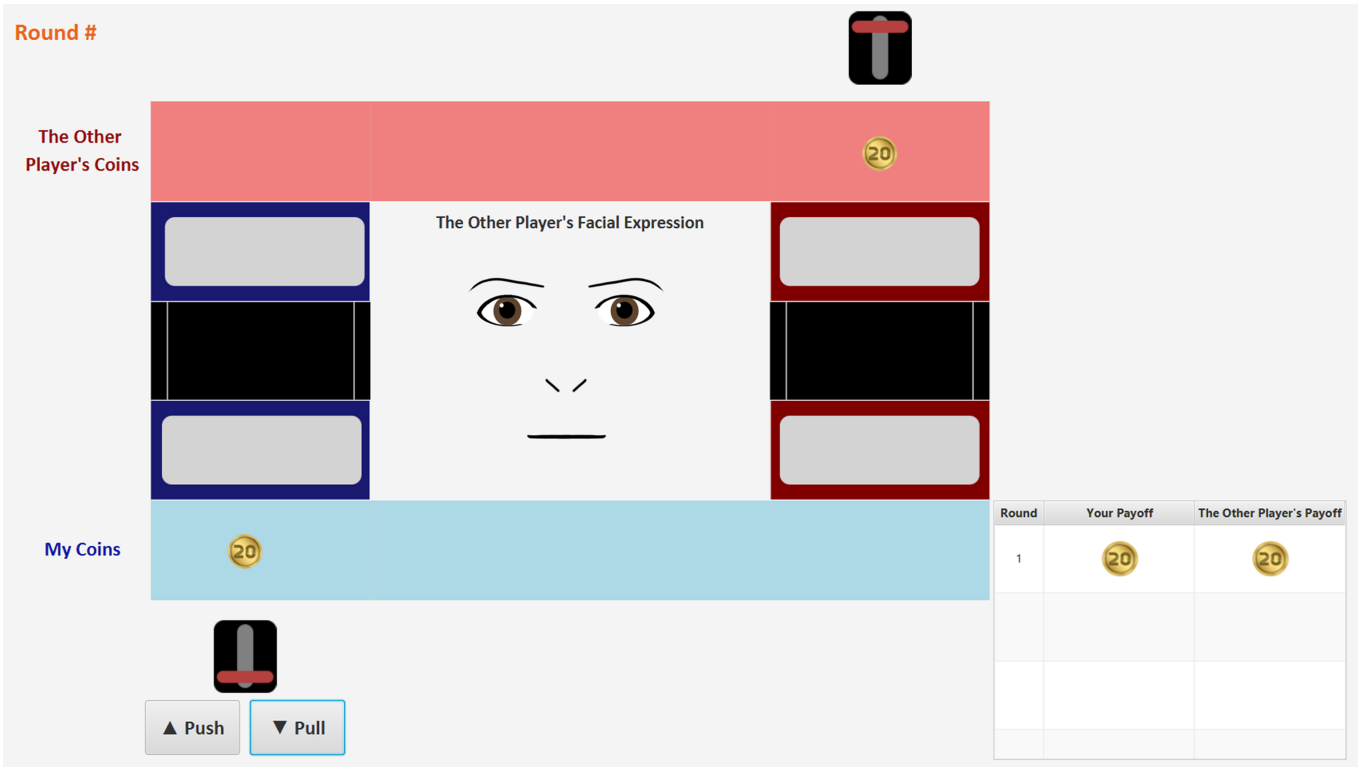


图 S4：图形用户界面：相互背叛。

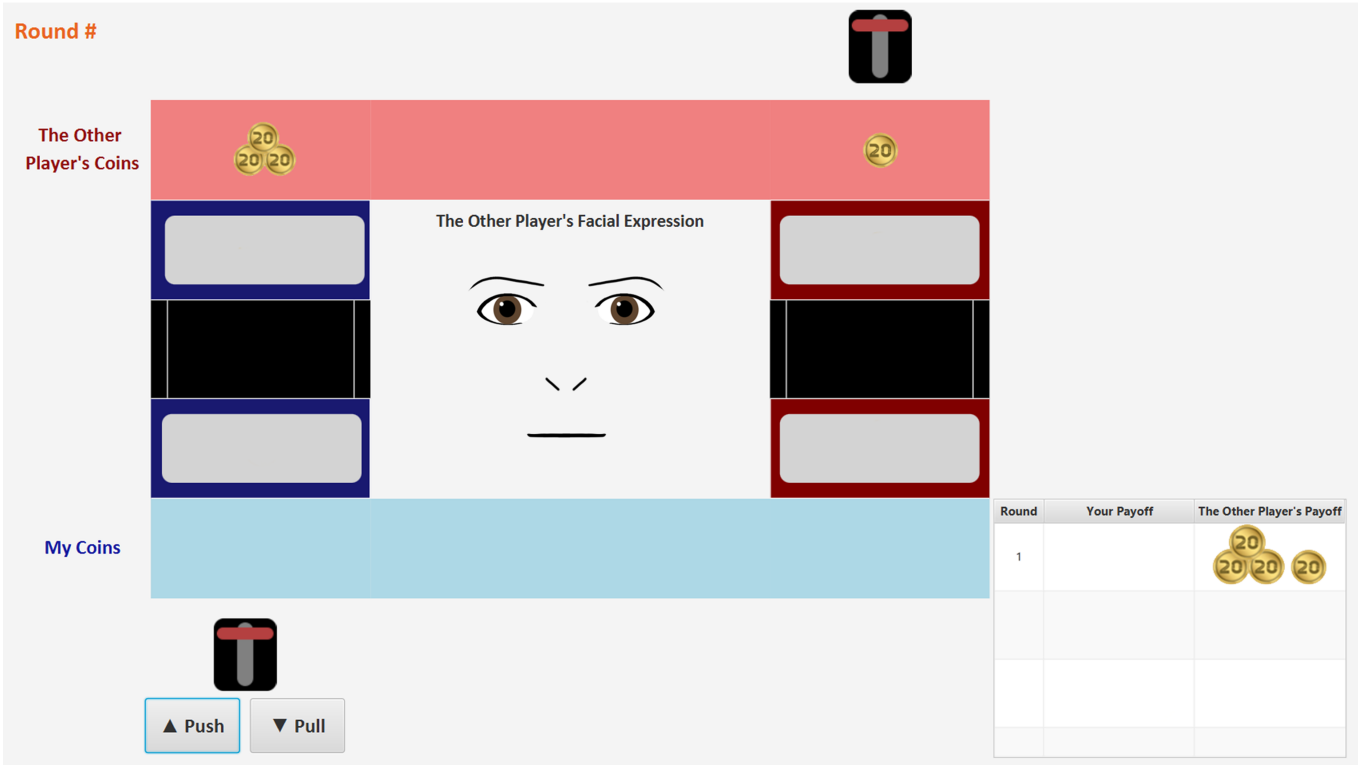


图 S5：图形用户界面：合作对抗背叛者。

生物特征采集与分类规范

使用 iMotions 生物特征研究平台（9.0 版）实时采集并分析参与者的面部表情。该平台集成了 Affectiva AFFDEX 情绪识别算法。AFFDEX 引擎采用基于深度学习的计算机视觉流程，检测面部标志点，并依据埃克曼和 Friesen 建立的面部动作编码系统（FACS）对表情进行分类 [1]。

分析分为三个阶段。首先，算法使用方向梯度直方图（HOG）和支持向量机（SVM）分类器，从视频流中分离出参与者的面部。其次，它映射 34 个不同的面部标志点（包括眼角、鼻尖和嘴部轮廓），以建立面部的几何构型。最后，分析这些标志点之间的关系，以检测特定的动作单元（AUs）——构成表情的基本单元（例如 AU12“嘴角上扬”）。随后，通过预训练的深度学习模型汇总这些 AUs，生成中性状态及七种核心情绪状态的概率分数（0–100）：喜悦、愤怒、惊讶、恐惧、轻蔑、厌恶和悲伤。

Affectiva AFFDEX 架构基于大规模自然情境数据集进行训练，该数据集包含数百万张面部帧，确保其在不同人口群体、光照条件和头部姿态下均具有很高的稳健性（McDuff 等人 [2]）。其分类准确率还通过独立研究与人类专家编码员的标注结果进行了进一步验证。St¨ockli 等人 [3] 报告称，对于明确且强度较高的不同表情，该软件的吻合率与人类编码员相当，尤其是喜悦表情（准确率 $> 90\%$）。为确保特定实验设置中的数据质量，我们维持了受控的光照条件，并固定参与者的位置，以尽量减少可能降低分类器性能的极端头部姿态变化。

图 S6 显示 iMotions 模块捕捉到中性表情。中性表情通常意味着没有效价，投入度较低（或没有投入度）。图 S7 显示该模块捕捉到喜悦表情。正向表情会带来正效价和高投入度；而当参与者注视屏幕时，负向表情则会带来负效价和高投入度。

显示和分析情绪

在面部表情交流处理中，参与者的面部表情会通过虚拟形象，实时动态地传达给与其配对的参与者。虚拟形象在任一时刻所显示的表情，都对应于强度值最高的情绪；每种情绪均以 0–100 的量表量化。系统设定了 60% 的阈值：若某一时刻没有任何情绪超过此值，则显示中性虚拟形象。如果软件未能检测到面部（例如，参与者将视线移开屏幕），则不会传达任何虚拟形象。

在事后回归分析中，我们使用原始的连续概率分数（0-100），而非经过阈值处理的二元状态。这样一来，我们便能检测到细微的情绪变化和“前情绪”（例如，喜悦程度从 15 上升到 50），这些变化表明参与者存在内在的情感投入，即使其强度尚不足以触发虚拟形象显示。

为最大限度地提高面部表情分析的准确性和效率，我们使用时间戳标注了每次交互中的关键事件。这些事件包括一轮次的开始、选择行动的时刻、观察联合动作的时刻、一轮次的结束，以及交互的终止。对于每一轮次，时间线被划分为两个区间：参与者选择行动之前的区间，以及选择行动之后的区间。我们排除了参与度低于30%的时间段。参与度反映了参与者对摄像头的视觉注意力；该摄像头位于用于查看信息和做出决策的显示器正上方，

用于查看信息和做出决策的显示器正上方。参与度较低表明参与者将视线移开了屏幕，或遮住了自己的面部。对于每一轮次，每种情绪均以两个均值进行概括：一个是测得的行动前强度值均值，另一个是测得的行动后强度值均值。

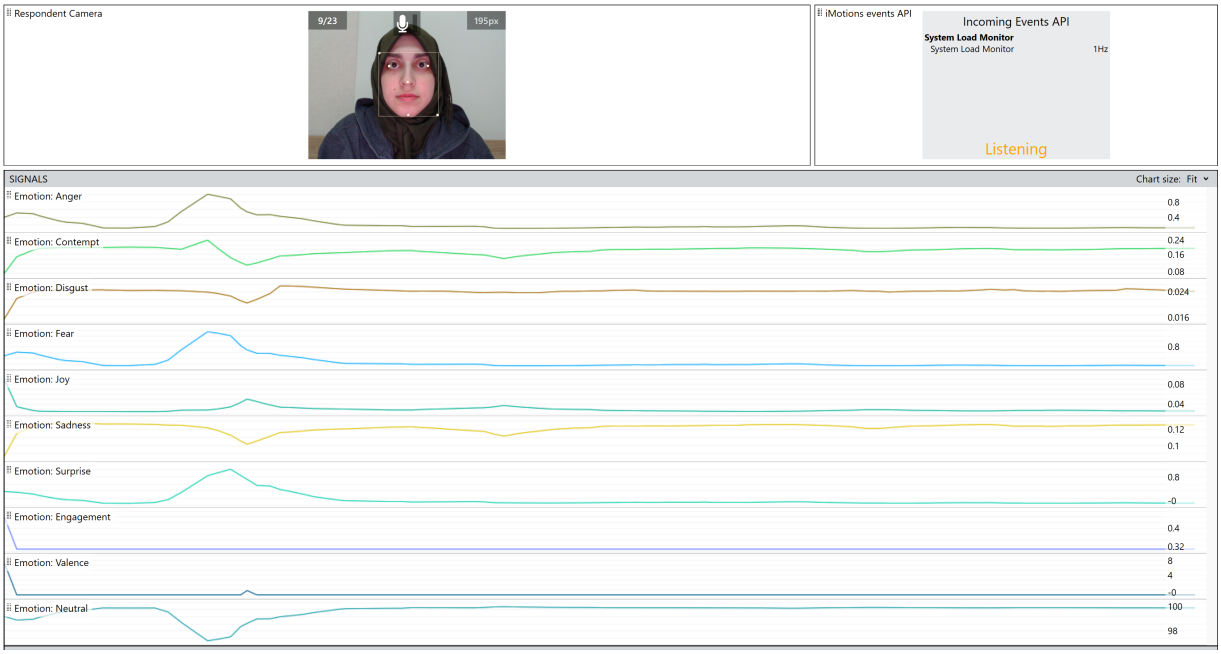


图 S6：iMotions 面部表情分析模块——中性表情。

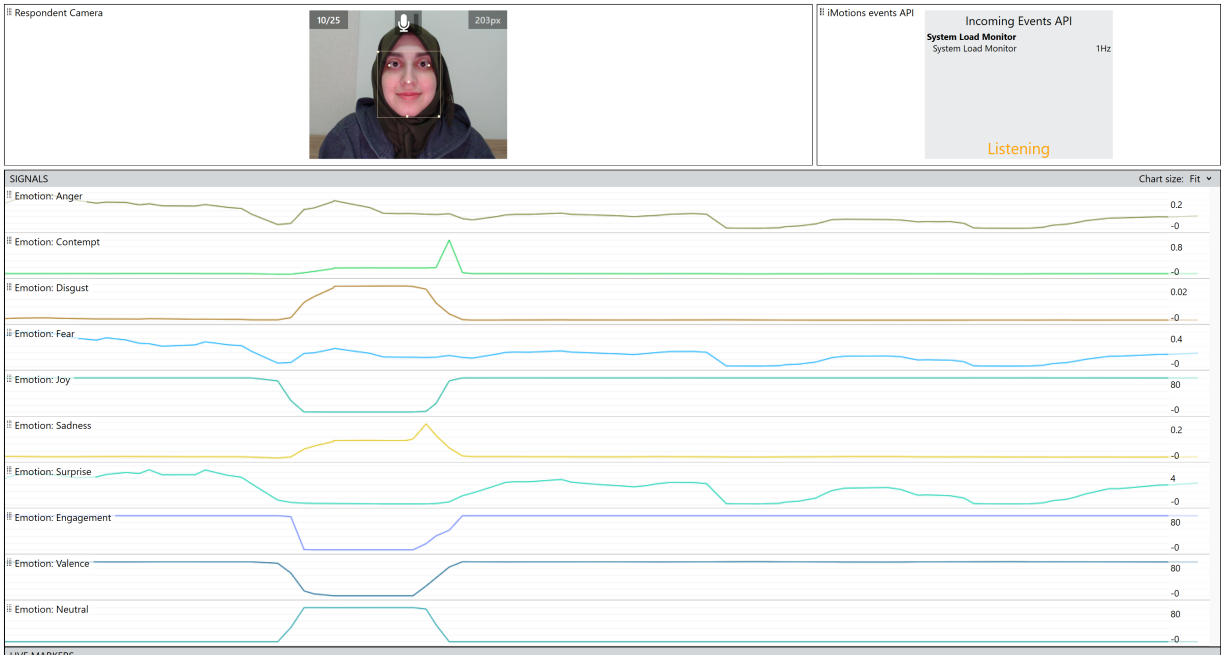


图 S7：iMotions 面部表情分析模块——愉悦表情。

虚拟形象验证与识别准确率

为确保我们的性别中立虚拟形象能够准确传达预期的情绪信号，我们在主要实验前开展了一项独立的验证研究。参与者（$N = 113$）观看了我们设计的虚拟形象，并被要求将其归入八种目标情绪类别之一。图 S8 展示了这些分类结果的混淆矩阵。对角线元素代表“命中率”（正确分类）。

对于我们的主要发现而言，至关重要的是，喜悦的识别率——它是与恢复合作相关的主要信号——极高（99%）。这证实，正文中观察到的“宽恕”机制是由明确无误的积极信号所驱动的。

与面部表情研究中的常见情况一致， [4, 5, 6], 对于细微程度较高的消极情绪，识别率较低，具体而言，轻蔑的识别率为 57%。然而，矩阵中的非对角线错误表明，误分类基本局限于其他负效价类别（例如，将轻蔑误认为厌恶或愤怒，或将其归为“难以分类”）。因此，尽管对消极情绪的具体标签可能有所不同，但其方向性信号（社会不满/负面情绪）仍得以保留。这表明，即使“轻蔑”这一具体标签并非总能被识别出来，虚拟形象仍成功传达了伴侣的不满。



图 S8：虚拟化身分类混淆矩阵

2 补充图

图 S9 至 S13 展示了参与者为评估设计的化身而填写的调查问卷。图 S14 至 S17 展示了参与者完成交互后填写的调查问卷。

化身评估调查

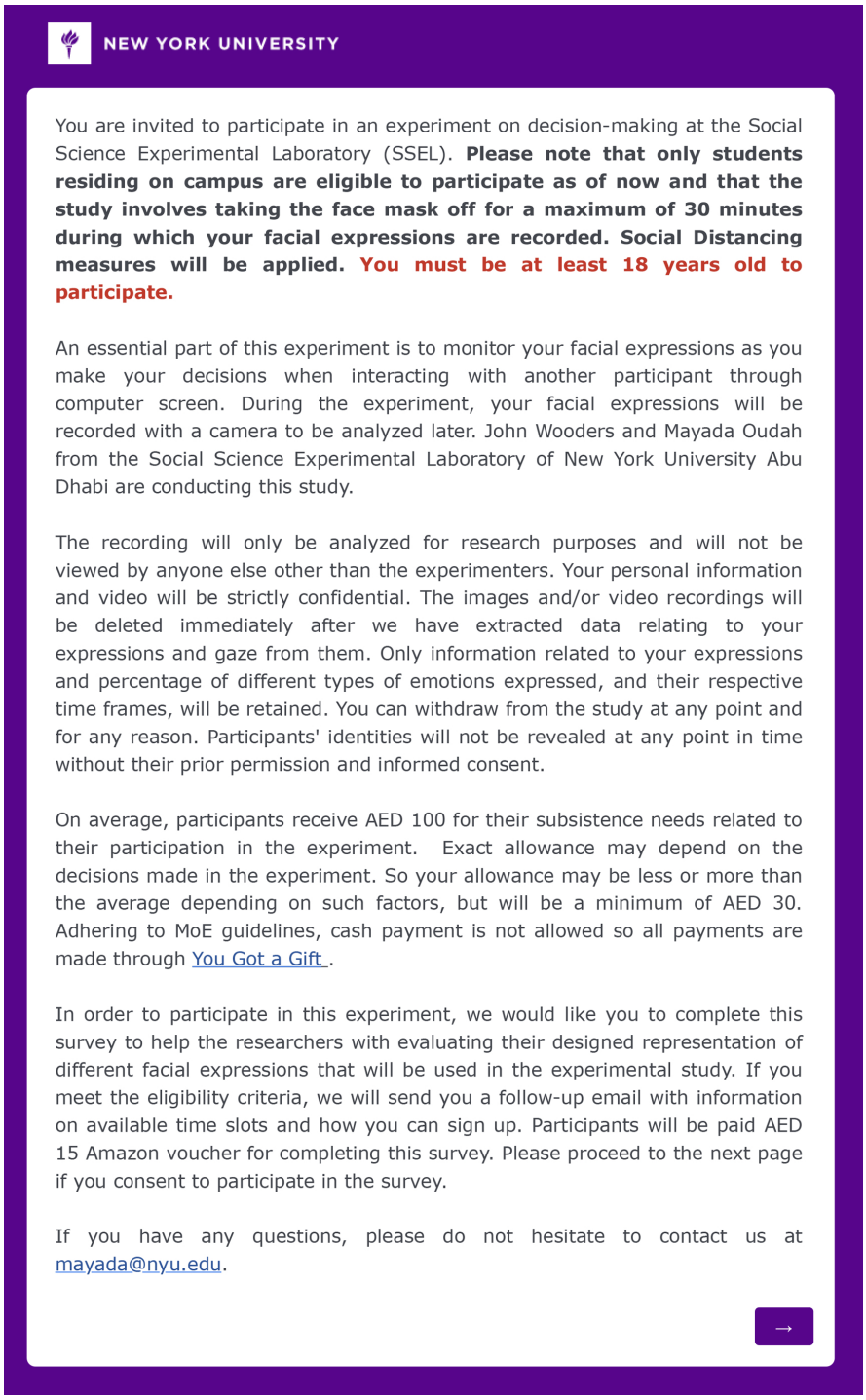


图 S9：调查问卷介绍。

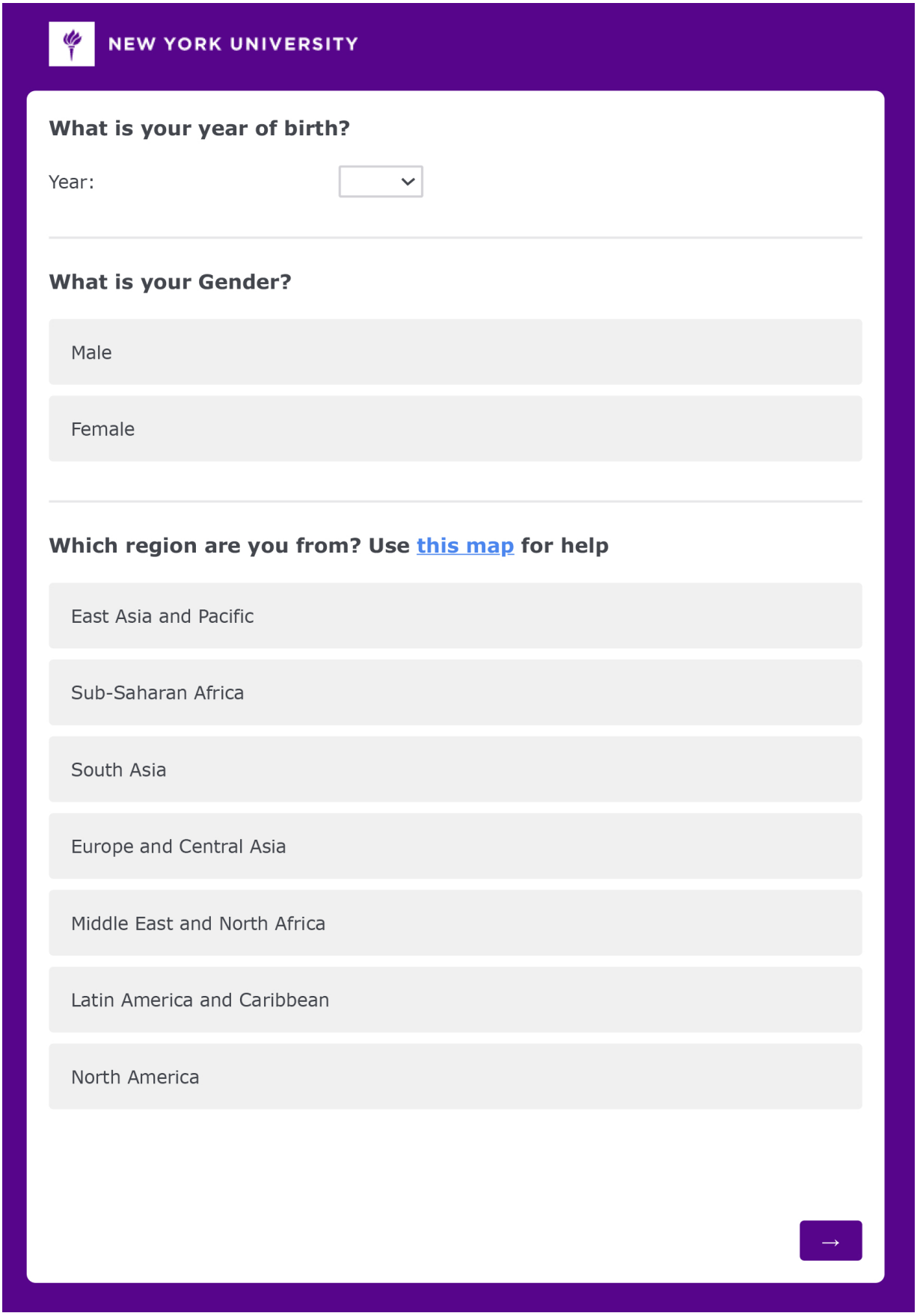


图 S10：人口统计学问题。

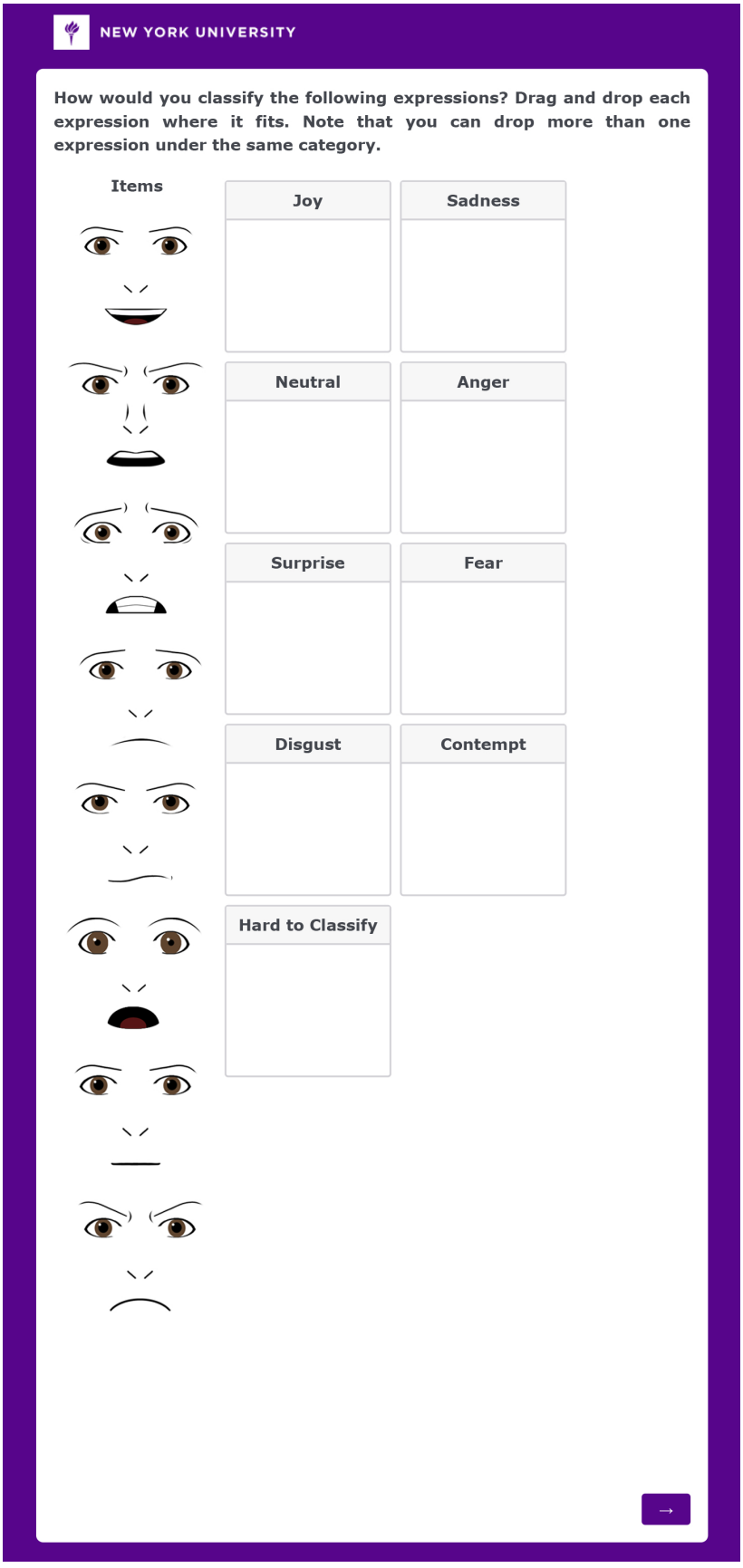

NEW YORK UNIVERSITY
How would you classify the following expressions? Drag and drop each expression where it fits. Note that you can drop more than one expression under the same category.
Items
Joy
Sadness
Neutral
Anger
Surprise
Fear
Disgust
Contempt
Hard to Classify
→

图 S11：头像分类。

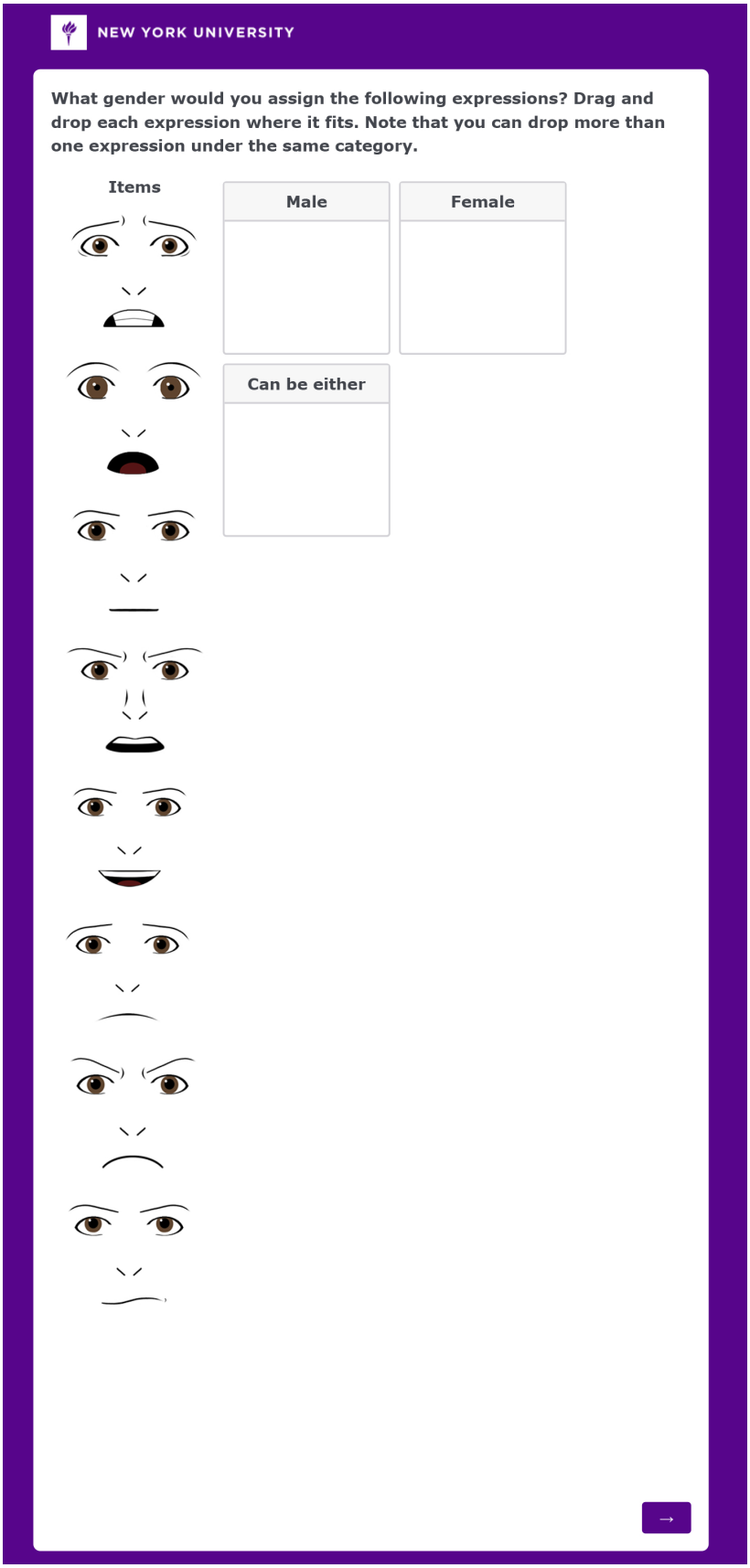


图 S12：虚拟形象性别分类。

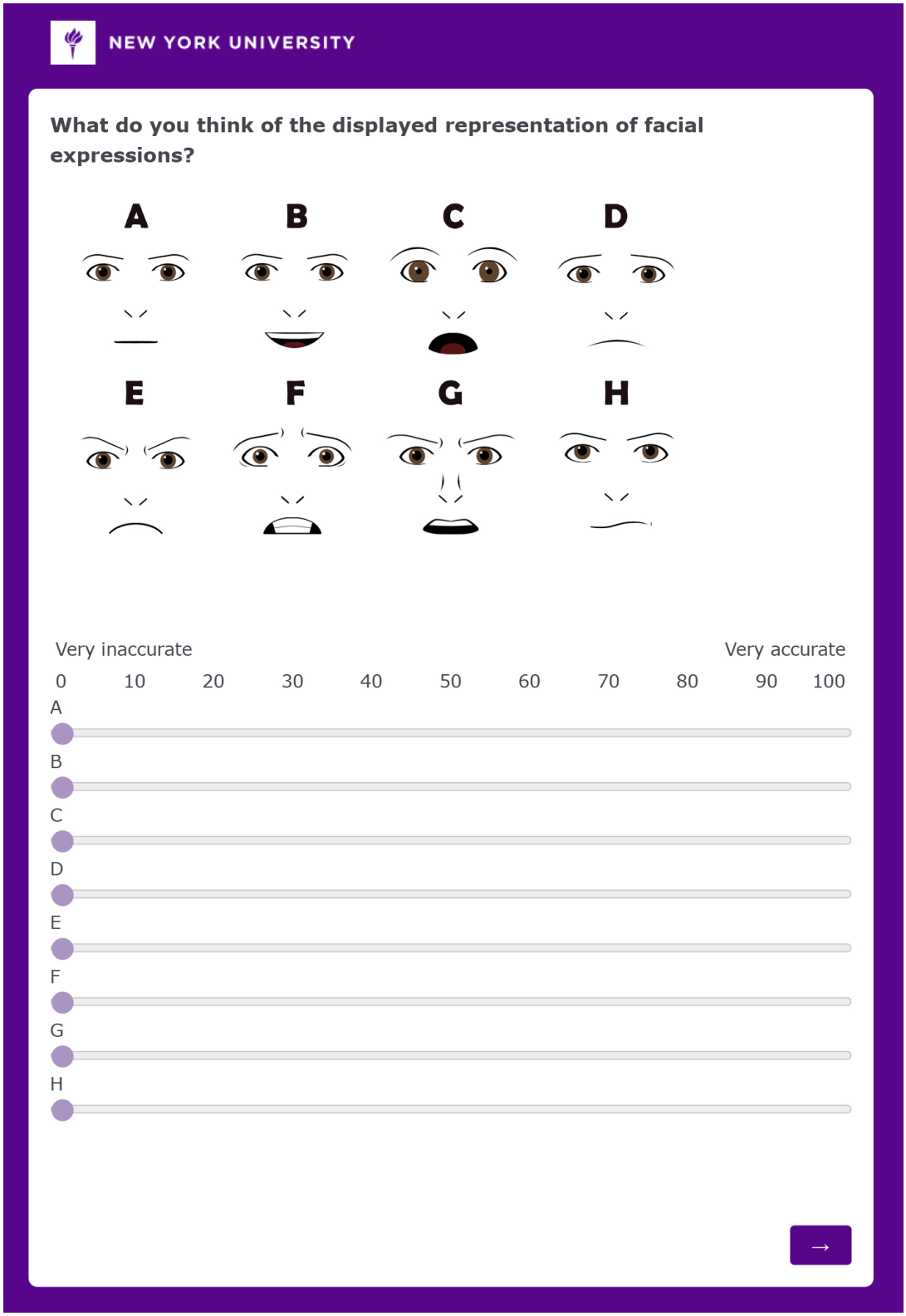


图 S13：头像准确性问题。

实验后调查

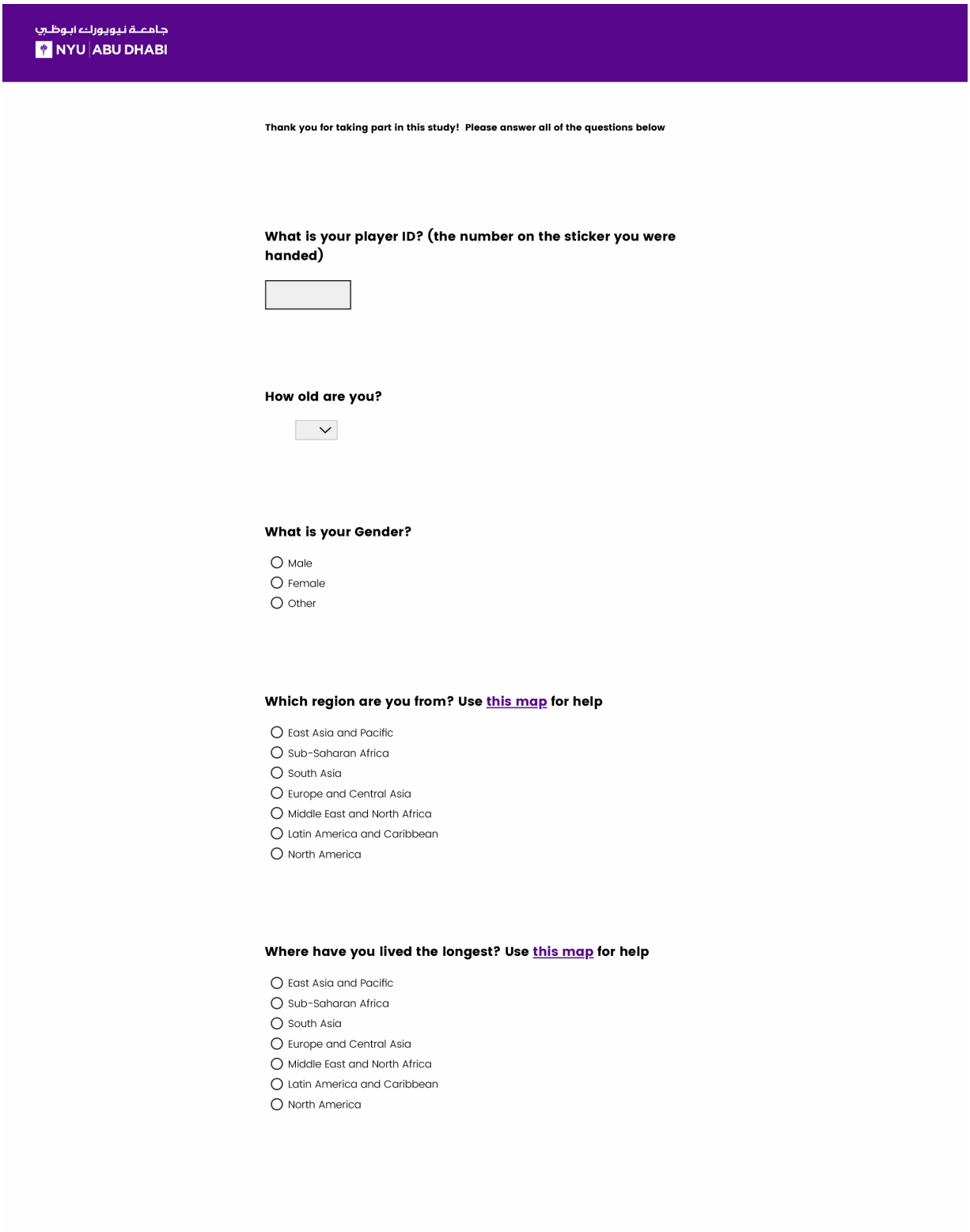


图 S14：实验后调查中的人口统计学问题。

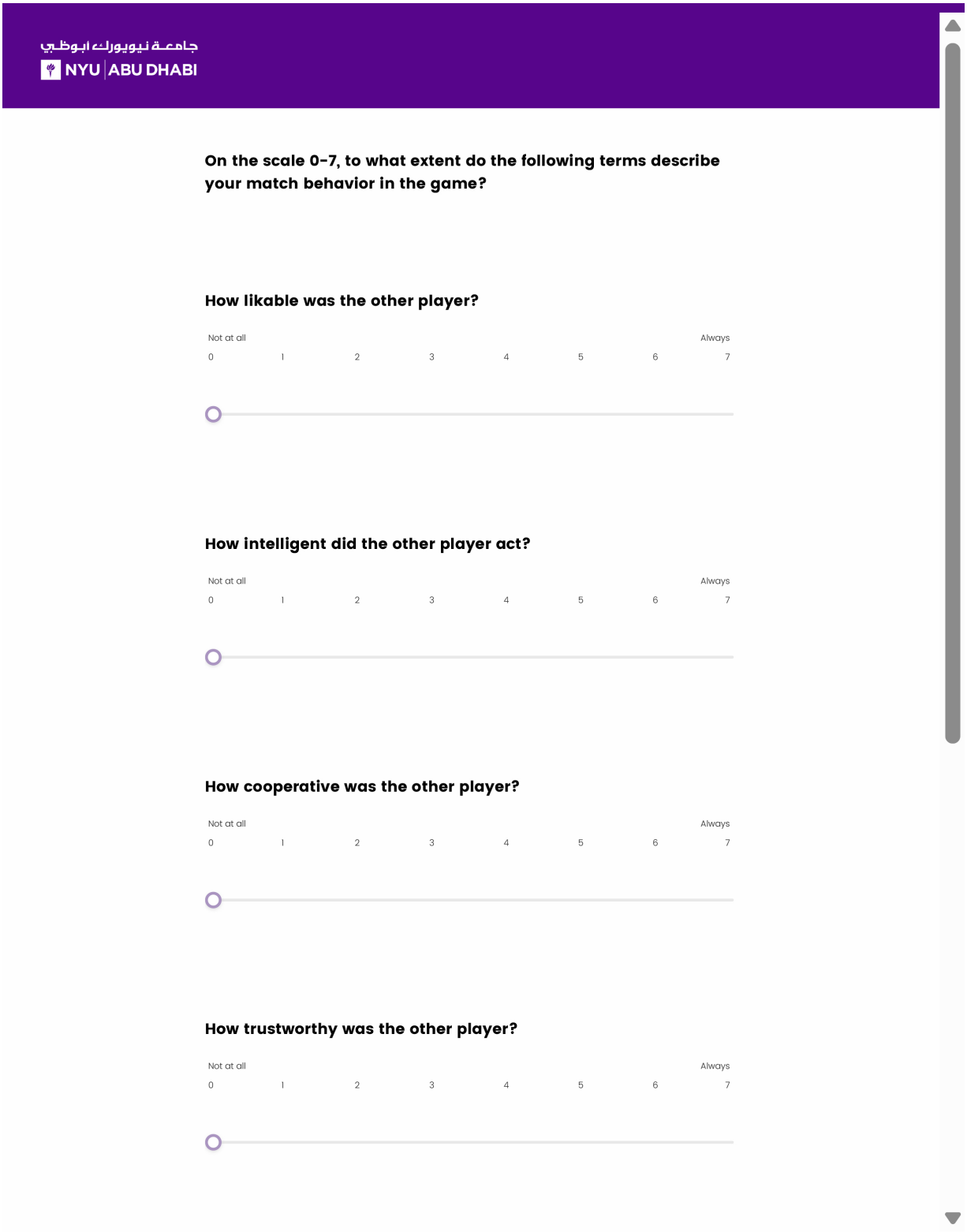


图 S15：对另一位玩家的可喜欢程度、智力、合作性和可信度进行评分。

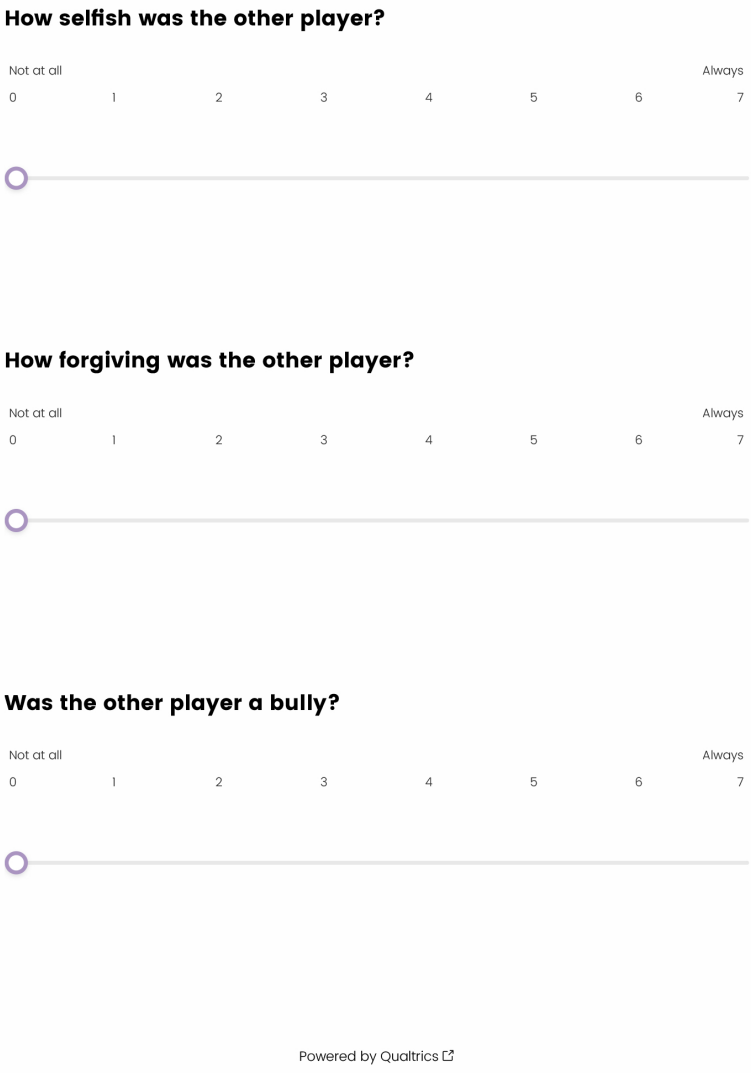


图 S16：对另一位玩家的自私、宽容程度及其是否为霸凌者进行评分。

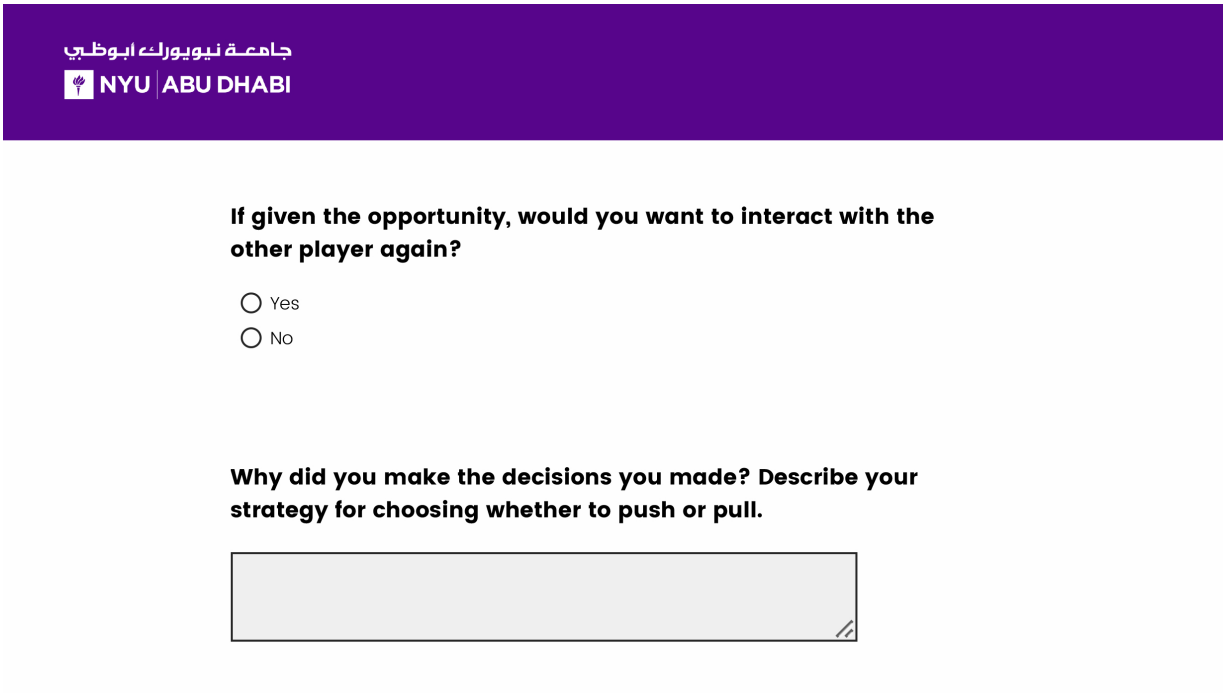


图 S17：“再次互动”问题以及对参与者决策策略的理解。

3 补充表

	合作
截距	0.69（0.16）***
处理（FC）	0.47 (0.07)***
AIC	5670.99
BIC	5690.63
对数似然	-2832.49
观测值数量	5160
组数：玩家ID	86
方差：玩家ID（截距）	2.04

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

表 S1：不同处理下合作的回归结果

	合作	合作	合作
截距	-2.00 (0.12)***	0.59 (0.16)***	-2.32 (0.91)*
处理（FC）	0.19 (0.09)*	0.27 (0.07)***	0.14 (0.10)
事前喜悦（自身）		0.64 (0.09)***	0.23 (0.65)
前次喜悦（他方）		0.21 (0.09)*	0.31 (0.68)
先前合作（己方）	1.69 (0.09)***		1.69 (0.09)***
先前合作（他方）	2.76 (0.09)***		2.74 (0.09)***
前次蔑视（己方）			-0.07 (0.68)
前测厌恶（自身）			-0.66（0.92）
前测恐惧（自身）			0.37（1.08）
前测中性（自身）			-0.04（0.65）
前测悲伤（自身）			0.40 (0.79)
前测惊讶（自身）			-1.08 (0.90)
前测轻蔑（他人）			0.48（0.72）
前测厌恶（他人）			-0.82 (0.99)
恐惧前（他人）			0.05（1.11）
中性前（他人）			0.36 (0.68)
悲伤前（其他）			0.84 (0.83)
惊讶前（其他）			-0.91 (0.88)
AIC	3580.65	5617.29	3589.35
BIC	3613.22	5650.03	3713.13
对数似然	-1785.32	-2803.64	-1775.67
观测值数量	4988	5160	4988
组数：玩家ID	86	86	86
方差：玩家ID（截距）	0.38	2.04	0.38

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

表 S2：按处理、先前行为和先前情绪划分的合作回归结果

参考文献

[1]Ekman, P. & Friesen, W. 跨文化面部与情绪中的恒常性。《人格与社会心理学杂志》**17**, 124–129 (1971).[2] McDuff, D. 等. Affdex 软件开发工具包：跨平台实时多面部表情识别工具包。载于《2016 年计算机协会人机交互会议扩展摘要集》，人因工程部分

s,

3723–3726（计算机协会，纽约，纽约州，美国，2016）。[3] St¨ockli, S., Schulte-Mecklenbeck, M., Borer, S. & Samson, A. C. 使用 Affdex 和 FACET 进行面部表情分析：一项验证研究。《行为研究方法》**50**, 1446–1460 (2018). URL https://api.semanticscholar.org/CorpusID:27426298.[4] Matsumoto, D. & Ekman, P. 蔑视表情、标签与描述之间的关系。《人格与社会心理学杂志》**87**, 529–40 (2004).[5] Russell, J. A. 蔑视表情与相对性论题。《动机与情绪》**15**, 149–168 (1991).[6] Ekman, P., O'Sullivan, M. & Matsumoto, D. 面部表情判断中的语境混淆：对"蔑视表情与相对性论题"的回应。《动机与情绪》**15**, 169–176 (1991).