混合通信中远程参与者的临场感建模及其在虚拟形象机器人行为设计中的应用

Takuma Miyaguchi[1] 和 Hideyoshi Yanagisawa[1*1] 工学研究生院，东京大学，日本东京

* 通讯作者

电子邮件：hide@mech.t.u-tokyo.ac.jp（HY）

摘要

面对面与在线混合交流存在这样的问题：只有本地参与者受到关注，而远程参与者被排除在外，或者本地参与者会忘记远程参与者。我们将这些问题视为缺乏远程参与者临场感的表现。据我们所知，作为对临场感的首个定量研究方法，我们采用贝叶斯框架对混合通信中远程参与者的临场感进行了建模。我们还将从贝叶斯模型模拟中获得的知识应用于虚拟形象机器人的干预行为，并通过虚拟形象机器人干预远程参与者的行为，鼓励本地参与者发言。我们假设，对远程参与者心理状态的信念是影响其临场感的一个因素。我们将临场感表述为对他人心理状态的模型证据。随后，我们采用主动推断框架对虚拟形象机器人的行为如何影响本地参与者的陈述进行建模，并将远程参与者的临场感纳入其中作为潜变量。通过使用该模型进行模拟，我们能够预测虚拟形象机器人的注视行为可以鼓励本地参与者发言。基于仿真结果，我们设计了虚拟形象机器人的注视行为。最后，我们考察了所设计的虚拟形象机器人注视行为的有效性。该注视行为表达了远程参与者对本地参与者更多的注意和兴趣，但本地参与者在会议任务中表达的意见却更少。结果表明，在实验任务情境下，注视行为增强了远程参与者的临场感，同时抑制了本地参与者发言。我们认为，临场感对参与者是否愿意表达意见有着相当大的影响。值得采用贝叶斯方法进一步探究临场感及其控制方法的影响

模型。

引言

近年来，许多会议和讲座采用结合面对面交流与在线交流的混合形式举行。研究报告了混合通信中存在的问题 [1]。具体而言，问题在于只有本地参与者受到关注，而远程参与者被排除在外，且本地参与者会忘记远程参与者。一些研究将混合通信中遇到的问题归因于远程参与者缺乏临场感，并尝试增强其临场感。例如，Yankelovich 等人开发了一种结合可远程控制的显示器与高品质扬声器的设备，以增强远程参与者的社会临场感 [1]。Van Dijk 等人提出了一种三维界面，通过调整摄像机和图像的布局，使画面能够自然呈现注视方向，并研究了其对社会临场感感知的影响 [2]。然而，这些研究仅限于定性讨论；据我们所知，尚未有人提出使用数学模型对临场感进行定量研究的方法。

该数学模型使我们能够全面预测潜变量变化对临场感的影响，以及临场感所产生的影响。这些预测可能超出人类的直觉。此外，我们可以沿着这一循环开展研究：首先，通过模拟预测潜变量集合的变化对临场感的影响，然后通过实验发现新的潜变量并更新模型。因此，我们认为，提出临场感的数学模型将有助于更有效地解决与临场感相关的问题。

Ogasawara 等人使用贝叶斯感知模型 [3]，构建了虚拟现实中其他虚拟化身的存在感模型。他们将存在感建模为虚拟化身时空定位的模型证据。简而言之，他们对虚拟化身的感知进行了物理建模。然而，我们构建了远程参与者的临场感模型，这种临场感会在混合通信中引发问题。通过探究这些问题的成因，我们提出了一个关于临场感影响因素的假设。

此外，我们利用临场感的数学模型，提出了一种能够促进本地参与者发表陈述的虚拟形象机器人行为。在具有不同地位层级的群体中，低地位成员往往不愿表达他们对人际风险的意见。我们使用虚拟形象机器人介入高地位

成员的行为，试图解决这一问题。

本研究有两个目标。首先，对混合通信中远程参与者的临场感进行形式化。其次，我们将贝叶斯模型模拟得出的结果应用于虚拟形象机器人的介入行为，并通过让虚拟形象机器人介入远程参与者的行为，促进本地参与者的陈述。

在远程人员临场感的数学建模中，我们首先说明与虚拟形象机器人进行的会话实验，以及关于临场感因素的假设。随后，我们介绍主动推断，并说明远程参与者临场感的数学建模。

在设计能够鼓励本地参与者在讨论中表达意见的虚拟形象机器人的注视行为时，我们阐述了用于设计促进本地参与者陈述的虚拟形象机器人行为的数学建模和模拟。重点关注了在具有不同社会地位层级的群体中发表的陈述。该模型将远程参与者的临场感视为潜变量。此外，还提出了一种基于模拟结果的虚拟形象机器人的注视行为。

在实验：虚拟形象机器人的注视行为干预对本地参与者陈述的影响，中，我们描述了一项实验，旨在考察所设计的注视行为的有效性。我们重现了一种情境：在社会地位层级不同的群体中，低地位成员不愿表达自己的意见。

方法

使用虚拟形象机器人的意义

本研究在混合通信中使用虚拟形象机器人作为远程参与设备。该虚拟形象机器人是索尼集团公司开发的原型机，我们正与该公司开展联合研究。图1展示了虚拟形象机器人的概况。虚拟形象机器人通过扬声器播放远程参与者的声音，还会同步呈现远程参与者的注视、眨眼和头部旋转。

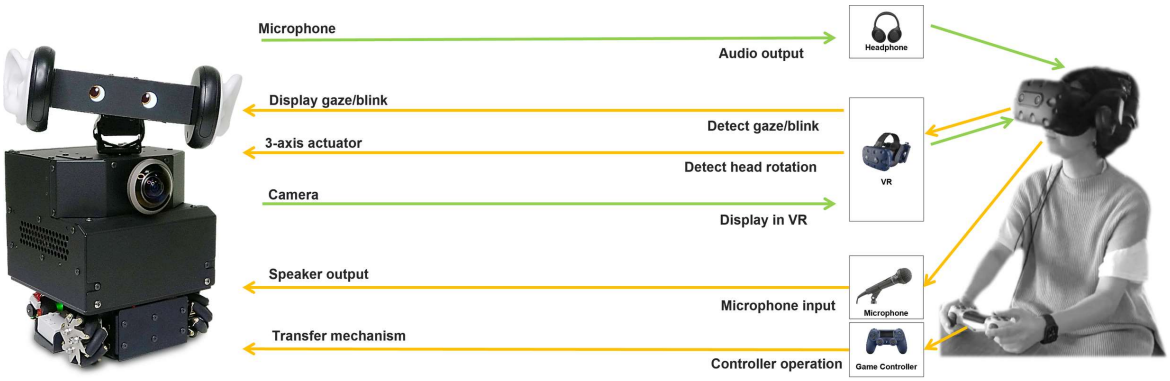


图1。本研究中使用的虚拟形象机器人。

这两项功能对于使用虚拟形象机器人至关重要。首先，虚拟形象机器人能够准确再现远程参与者的注视方向。在使用平面显示器的混合系统中，很难判断远程参与者的注视方向。据报道，与在显示屏上呈现远程参与者的面部相比，旋转机器人的三维头部能够更准确地传达注视方向[4]。虚拟形象机器人会随着远程参与者的头部旋转而转动头部。我们认为，虚拟形象机器人能够准确再现注视方向。

其次，虚拟形象机器人可以介入远程参与者非言语行为（即头部旋转、注视和眨眼）的传输。虚拟形象机器人不仅反映远程参与者的行为，还会对其进行编辑、转换或自动化处理。非语言行为表达情绪[5] 和认知状态[6]。我们尝试通过介入非语言行为并控制情绪和认知状态的传输，来调控临场感。

远程人员临场感的数学建模

利用虚拟形象机器人进行会话实验，以提出有关临场感定义及其影响因素的假设。

我们将使用虚拟形象机器人进行会话实验，作为临场感数学建模的第一步。实验旨在提出有关远程参与者临场感的定义及其影响因素的假设。

一名远程参与者和三名本地参与者参与了一项会话任务。图2展示了会话实验中参与者的布置情况。两名

将对话形式（演示和讨论）设定为任务。这样做的原因是，有必要观察或体验不同的对话场景（即一名参与者与其他人交谈、参与者之间自由交替发言，以及一名参与者向另一名参与者讲话的场景）。

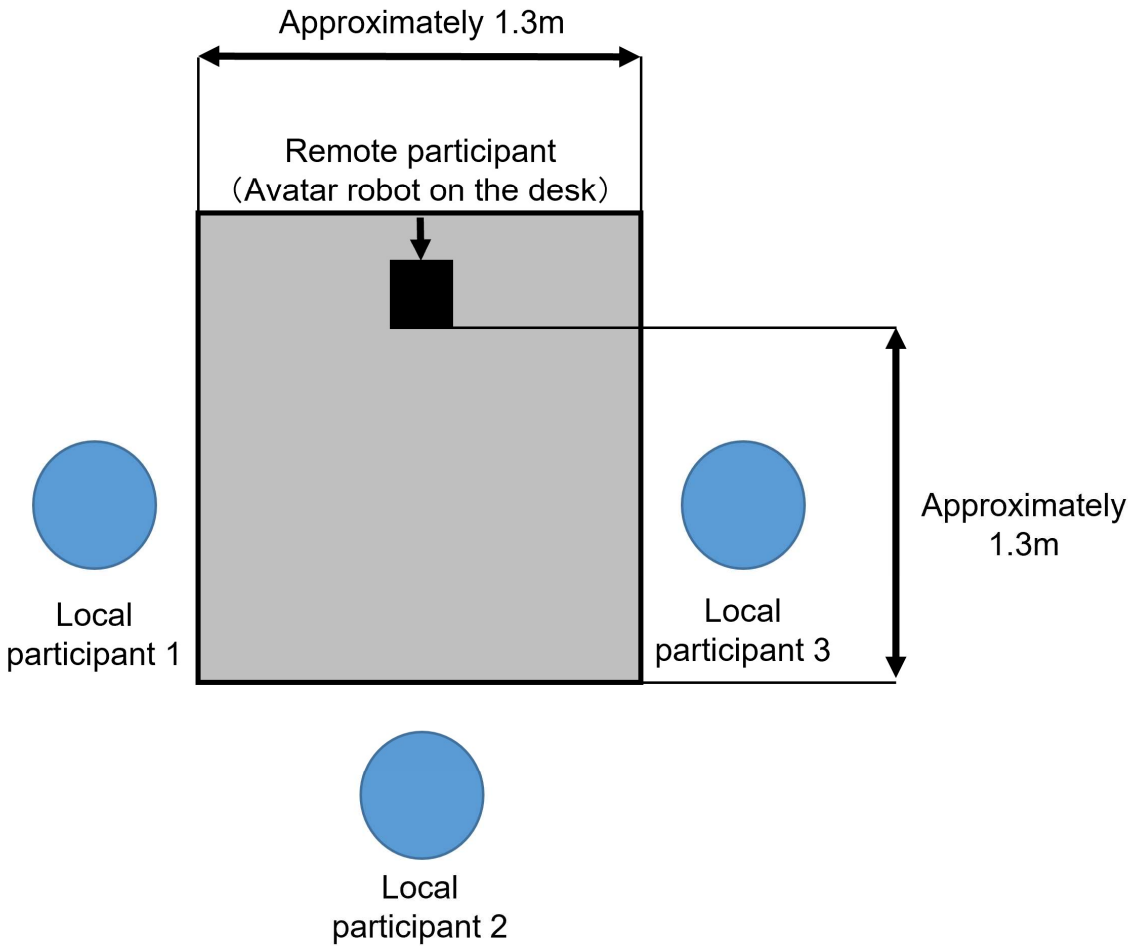


图 2. 对话实验中参与者的布置。

所研究的临场感问题归因于混合通信中的问题，例如只有本地参与者受到激励、远程参与者被排除在外，或本地参与者忘记了远程参与者。因此，我们将讨论只有本地参与者变得兴奋、远程参与者被排除在外，或本地参与者忘记远程参与者的原因。以下列举了这些现象的成因：

- 由于只有远程参与者会经历延迟，因此他们在积极轮换时很难轮流发言。

- 本地参与者很难理解远程参与者的心理状态，例如他们的想法和认知状态。

- 本地参与者很难让远程参与者轮流发言，因为他们无法预测远程参与者的回应。

在这三点中，我们重点关注了后两点，它们与对远程参与者的识别有关。虚拟形象机器人可以显示远程参与者的注视、眨眼和头部旋转，但无法显示其他行为，例如面部表情和手势。通过显示屏呈现远程参与者的混合系统可以展现其面部表情和手势，但很难知道远程参与者正在看向何处。在任何混合系统中，能够提示远程参与者心理状态的线索都少于面对面交流。因此，很难判断远程参与者的心理状态。此外，当远程参与者的心理状态不明时，也很难预测他们的回应。交流错误的风险，例如对尚未准备好回应的远程参与者说话，以及难以根据远程参与者的想法进行交流，似乎是难以与远程参与者交谈的原因。

我们假设，难以检测远程参与者的心理状态，与混合通信中的问题以及远程参与者的临场感有关。我们还假设，对远程参与者心理状态的信念会影响远程参与者的临场感。在本研究中，我们采用主动推断框架[7]，以数学方式刻画对远程参与者心理状态的信念。

主动推断

本节将介绍用于建模临场感的主动推断方法。详细说明见[7]。

主动推断是一种数学框架，其核心思想是：感知和学习可以理解为最小化变分自由能（VFE），而动作选择和规划则可理解为最小化期望自由能（EFE）[7]。为了解释主动推断，首先介绍贝叶斯推断。

贝叶斯推断利用贝叶斯定理，根据外部世界隐藏状态所产生的观测来推断这些隐藏状态。这里的“隐藏”是指状态本身无法被直接观测，只能通过

观测（感觉刺激）来推断。式（1）展示了贝叶斯定理。

$$p(s|o) = \frac{p(o|s)p(s)}{p(o)} \tag{1}$$

隐藏状态记为 s，观测记为 o。术语$p(s|o)$指通过推断获得的、编码了对状态信念的后验。术语$p(o|s)$指似然，它编码了有关观测与状态之间映射的知识。术语 $p(s)$指先验，它编码了观测发生之前对状态的预测。术语 $p(o)$指模型证据，它编码了生成模型的合理性。贝叶斯定理指出，似然与先验相结合可推导出后验。

然而，除了最简单的分布之外，贝叶斯定理都无法直接计算。这是因为计算模型证据需要对生成模型所有可能状态下观测的概率求和。公式（2）展示了模型证据的计算公式：

$$p(o) = \sum_s p(o,s) = \sum_s p(o|s)p(s) \tag{2}$$

因此，采用了变分推断来近似后验。引入了近似后验 $q(s)$。随后计算变分自由能，并将其表示为 $q(s)$与生成模型$p(o,s)$之间的库尔贝克-莱布勒散度（KL散度）。KL散度是一种用于衡量两个分布之间差异的指标。公式（3）展示了变分自由能。

$$F = \sum_s q(s) \ln \frac{q(s)}{p(o,s)} \tag{3}$$

对 $q(s)$ 进行变分更新。如果发现 $q(s)$ 使变分自由能达到最小值，则将该点处的后验 $p(s|o)$ 近似为真实后验。

这解释了主动推断中的感知机制。接下来将讨论动作选择。

在动作选择过程中，大脑还会做出能够带来未来观测的决策，以最小化变分自由能。然而，未来结果尚未被观测到。因此，我们计算并最小化了期望自由能（即变分自由能的期望值）。动作策略 π 的期望自由能表示为公式（4）[7]：

$$G_\pi = \sum_{s,o} q(o,s|\pi) \ln \frac{q(s|\pi)}{p(o,s|\pi)} \tag{4}$$

$$= \sum_{s,o} q(o,s|\pi) \ln \frac{q(s|\pi)}{p(s|o,\pi)} - \sum_{o} q(o|\pi) \ln p(o|\pi)$$

$$\approx \sum_{s,o} q(o,s|\pi) \ln \frac{q(s|\pi)}{q(s|o,\pi)} - \sum_{o} q(o|\pi) \ln p(o|C)$$

公式（4）的第三行进行了两处替换。首先，用近似后验 $q(o,s|\pi)$ 替代真实后验 $p(s|o,\pi)$。其次，在第二项中，用 $p(o|C)$ 替代$p(o|\pi)$。C 是先验偏好，而 $p(o|C)$ 编码了智能体根据先验偏好所偏好的观测。

基于主动推断框架的临场感数学建模

使用主动推断对混合通信中远程参与者的临场感进行建模。

Ogasawara 等人将虚拟现实中感知与虚拟化身时空定位相关的刺激时，对另一个人的虚拟化身存在感的感受建模为模型证据的大小，或惊异度的低下程度 [3]。我们假设，对远程参与者心理状态的信念是对话实验中产生远程参与者临场感的一个因素。通过考虑对远程参与者心理状态的信念来感受临场感，并采用 Ogasawara 等人提出的模型进行建模。

混合通信中远程参与者的临场感被建模为远程参与者心理状态的模型证据。关于远程参与者心理状态的大量模型证据表明其清晰度较高。因此，模型证据与远程参与者的临场感相关。借助该模型，可以将远程参与者的临场感表示为主动推断框架中的潜变量。

设计能够鼓励本地参与者在讨论中表达意见的虚拟形象机器人的注视行为

本节设计了虚拟形象机器人的行为，以鼓励本地参与者在讨论中表达意见，并解决不同社会地位层级的群体中表达意见的问题。为此，

我们对高地位成员的注视行为对低地位成员陈述的影响进行了建模，其中包括在主动推断框架中建模的临场感。随后，我们利用该模型进行了模拟，并获得了设计虚拟形象机器人行为的启示。

不同社会地位群体中表达意见的问题

在小组讨论中，信息共享和交换意见十分重要，因为它们能够提升群体绩效。信息共享能够让知识或认知水平存在差异的群体汇集并总结所有信息 [8]。据报道，群体中的不一致与创造力提升 [9] 及更好的决策 [10] 有关。

然而，为了共享信息或表达异议而发表意见，往往伴随着人际风险。在群体中反对主流意见可能导致被群体排斥 [11]。表达意见可能招致批评性反馈，并使他人觉得自己能力不足。如果群体中有人能够影响自己的晋升或薪资，那么表达意见就可能成为一项更为具体且风险更高的行动 [12]。因此，在地位层级不同的群体中，低地位成员会因担心人际风险而犹豫是否表达意见。

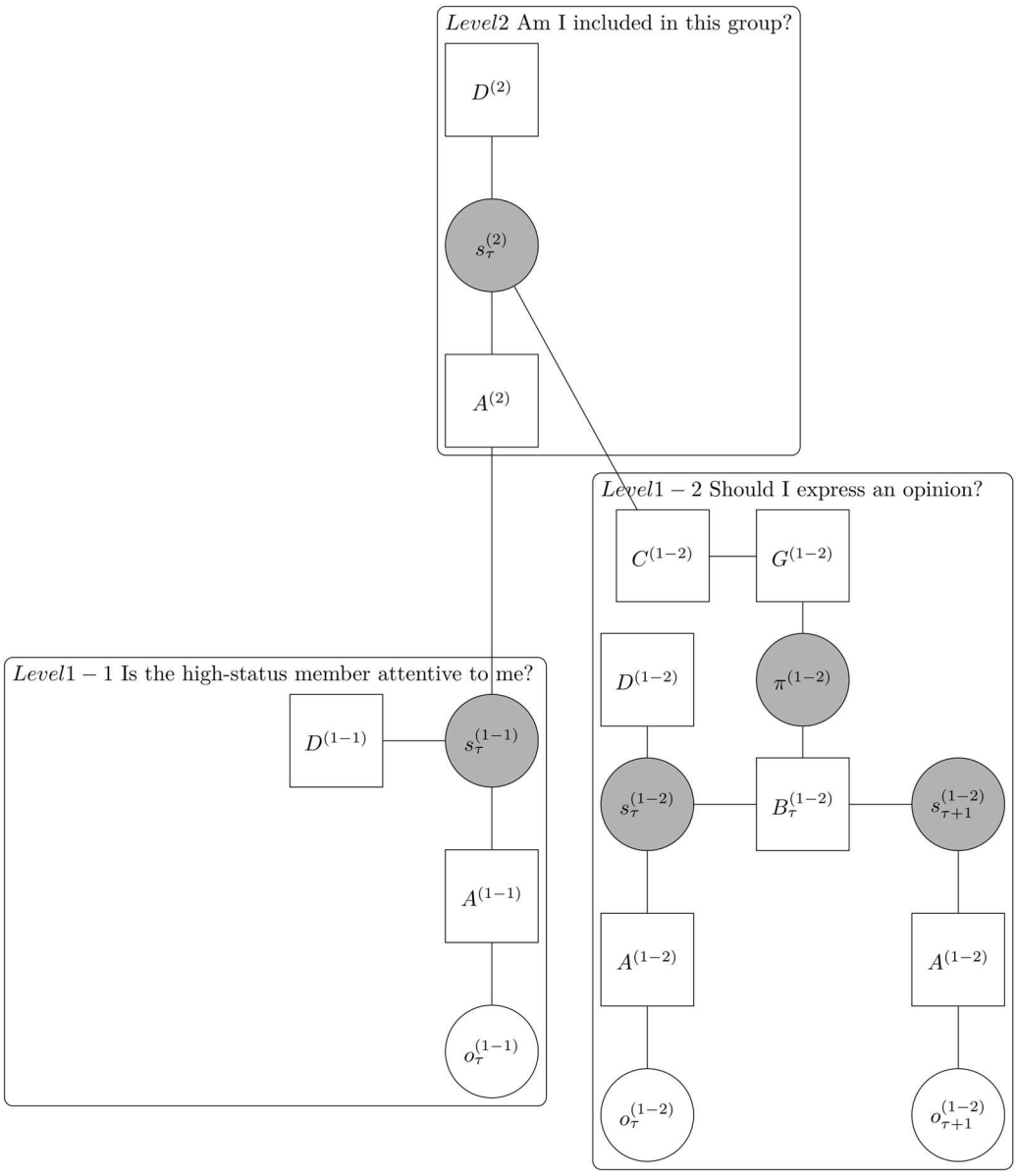
高地位成员的注视行为对低地位成员陈述的影响：数学建模

高地位成员的特征和行为会影响他人的陈述。据报告，团队领导者是否具有合作精神或教练型风格，是预测团队心理安全感的一个因素 [12]。心理安全感是团队成员共同持有的一种信念，即团队在人际关系中可以安全地承担风险；它与团队成员的学习行为有关 [12]。Shim 等人发现，当群体领导者更多地将注视投向一名低地位成员时，群体成员之间影响力的差异会缩小。[13]。

我们通过虚拟形象机器人干预高地位成员的非语言行为，从而提升低地位成员的陈述。为了设计能够促进低地位成员陈述的虚拟形象机器人行为，我们对高地位成员的行为对低地位成员陈述的影响进行了建模

，采用了主动推断框架。

图3展示了所提出的贝叶斯模型，该模型描述高地位成员的行为对低地位成员陈述的影响。s为隐藏状态；o为观测；A为状态与观测之间的似然映射（观测模型）；B为转移矩阵；C为先验偏好；D为初始状态先验。该模型分为三个字段：第1-1层、第2层级和1-2级。在所有字段中，推断主体都是低地位成员。各字段描述如下：



<escape>图3. 高地位成员的注视行为对低地位成员陈述的影响的贝叶斯模型.
</escape>

A：状态与观测之间的似然映射（观测模型），B：转移矩阵，C：先验偏好，D：初始状态先验，G：期望自由能，s：状态，o：观测，π：动作后验分布。上标表示变量所属的字段。下标表示时间步。在所有字段中，推断主体都是低地位成员。

在第1-1层中，低地位成员根据高地位成员的注视行为，推断其是否在注意。人类可以从他人的注视中判断对方在注意什么[14]。因此，接收到直接注视的智能体会推断对方正在注意该智能体。或者，接收到对方回避注视的智能体会推断对方没有在注意该智能体。观测模型 $p\left(o^{(1-1)}\middle|s^{(1-1)}\right)$ 由公式（5）描述。 ζ_{1-1} 是观测模型第1-1层的精度。 $\sigma()$是Softmax函数。添加$\mathrm{Exp}(-4)$以避免出现$\ln(0)$。通过变换公式（5）中的第3行和第4行，可以将与观测相关的噪声以及观测与状态之间映射的模糊性纳入考虑。

$$o^{(1-1)} = \{direct\ gaze \quad averted\ gaze\},$$
$$s^{(1-1)} = \{attentive \quad unconcerned\}$$
$$A_{1-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$A^{(1-1)}: p\left(o^{(1-1)}\middle|s^{(1-1)}\right) = \sigma(\zeta_{1-1} * \ln(A_{1-1} + \exp(-4))) \tag{5}$$

智能体在第1-1层推断高地位成员的心理状态。因此，第1-1层推断的模型证据对应于高地位成员的临场感。

在第2层级，低地位成员根据第1-1层推断的结果，推断自己是否被群体接纳。被接纳的状态对应较高的心理安全感。Hirak 等人 [15] 报告称，领导者的包容性与其他成员感知到的心理安全感呈正相关。领导者的包容性定义为领导者在与其他成员互动时表现出的开放性和可接近性。假设智能体会根据高地位成员群体的注意力和兴趣，认为自己被群体接纳。相反，假设智能体会认为高地位成员对自己的漠不关心具有排斥性。观测

模型 $p\left(s^{(1-1)}\middle|s^{(2)}\right)$ 由公式（6）描述。

$$s^{(1-1)} = \{attentive \quad unconcerned\},$$
$$s^{(2)} = \{included \quad excluded\}$$
$$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$A^{(2)}: p\left(s^{(1-1)}\middle|s^{(2)}\right) = \sigma(\zeta_2 * \ln(A_2 + \exp(-4))) \tag{6}$$

在1-2级，低地位成员选择自己的动作：表达意见或不表达。表达意见会引发群体反馈。我们假设智能体会学习群体是否认同自己的意见。观测模型 $p\left(o^{(1-2)}\middle|s^{(1-2)}\right)$ 如公式（7）所示。考虑到错误推断的概率时，观测模型会发生变化。

$$o^{(1-2)} = \{agreement \quad disagreement\},$$
$$s^{(1-2)} = \{agreement \quad disagreement\}$$
$$A^{(1-2)}: p\left(o^{(1-2)}\middle|s^{(1-2)}\right) = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \tag{7}$$

在1-2级，智能体对与动作选择相关的未来观测具有先验偏好。先验偏好编码了从观测中获得的奖励和厌恶。表达意见涉及人际风险。表达异议可能导致被群体排斥 [11]。我们为观测 $disagreement$设定了负偏好。考虑到心理安全感能够促进承担人际风险的学习行为 [12], ，我们将负偏好的变化幅度设定为由在第2层级推断出的群体包容性决定。

可以将先验偏好建模为取决于情境。在这种情况下，期望自由能由公式（8）[16]描述：

$$G_\pi^{(1-2)} = -\sum_{s^{(1-2)},o^{(1-2)}} q\left(o^{(1-2)}, s^{(1-2)}\middle|\pi^{(1-2)}\right) \ln \frac{q\left(s^{(1-2)}\middle|o^{(1-2)},\pi^{(1-2)}\right)}{q\left(s^{(1-2)}\middle|\pi^{(1-2)}\right)}$$
$$-\sum_{o^{(1-2)}} q\left(o^{(1-2)}\middle|\pi^{(1-2)}\right) \ln p\left(o^{(1-2)}\middle|s_\tau^{(2)}\right) \tag{8}$$

式（8）中的第二项 $p\left(o^{(1-2)}\middle|s_\tau^{(2)}\right)$ 是先验偏好，它取决于时间 τ的情境。当智能体认为自己属于群体且心理安全感较高时，智能体会认为承担人际风险是安全的。相反，当智能体认为自己被群体排除且心理安全感较低时，智能体会认为承担人际风险存在危险。

因此，只有当 $s^{(2)}=excluded$时，才会为 $o^{(1-2)}=disagreement$ 设定负偏好。先验偏好被描述为依赖语境的，如式（9）所示。

$$o^{(1-2)}=\{agreement \quad disagreement\},$$
$$s^{(2)}=\{included \quad excluded\}$$
$$p\left(o^{(1-2)}\middle|s^{(2)}\right)=\begin{bmatrix}0 & 0\\ 0 & -1\end{bmatrix} \tag{9}$$

式（8）中的期望自由能允许将先验偏好视为离散语境的函数。在对心理安全感水平对动作选择的影响进行建模时，我们按照式（10）所示构建了先验偏好。

$$C^{(2)}: p\left(o^{(1-2)}\middle|C\right)=p\left(s^{(2)}=included\right)p\left(o^{(1-2)}\middle|s^{(2)}=included\right)$$
$$+p\left(s^{(2)}=excluded\right)p\left(o^{(1-2)}\middle|s^{(2)}=excluded\right) \tag{10}$$

随后，式（11）对期望自由能进行了描述。

$$G_{\pi}^{(1-2)}=-\sum_{s^{(1-2)},o^{(1-2)}} q\left(o^{(1-2)},s^{(1-2)}\middle|\pi^{(1-2)}\right)\ln\frac{q\left(s^{(1-2)}\middle|o^{(1-2)},\pi^{(1-2)}\right)}{q\left(s^{(1-2)}\middle|\pi^{(1-2)}\right)}$$
$$-\sum_{o^{(1-2)}} q\left(o^{(1-2)}\middle|\pi^{(1-2)}\right)\ln p\left(o^{(1-2)}\middle|C\right) \tag{11}$$

基于贝叶斯模型的模拟

我们使用所提出的模型进行模拟，以研究高地位成员的注视行为对低地位成员陈述的影响，并据此设计虚拟形象机器人的干预行为。

我们对每种条件下低地位成员的动作选择概率进行了模拟。动作选择概率使用式（12）计算。

$$\pi=\{express\ an\ opinion \quad do\ not\ expree\ an\ opinion\}$$
$$p(\pi)=\sigma(-\gamma G) \tag{12}$$

术语 γ 表示期望自由能的期望精度。其初始值设为 1。术语 G 是由两个行动的期望自由能构成的向量。术语 $\sigma()$ 表示 Softmax函数。我们使用 [7] 中的 MATLAB 代码 spm_MDP_VB_X_tutorial.m 进行模拟。

虚拟形象机器人可以干预第1-1层中的观测（$o^{(1-1)}$）、第1-1层中状态与观测之间的似然映射（$p\left(o^{(1-1)}\middle|s^{(1-1)}\right)$），以及第1-1层中的先验（$p(s^{(1-1)})$），这些都表示观测前的预测。我们模拟了改变模型变量的条件。模拟条件列于表

1. 通过观察后续遵循期望自由能（EFE）的动作选择，我们将期望自由能（EFE）的期望精度（γ）初始值设为1。

表 1。模拟条件。

条件	观测 $o^{(1-1)}$	似然映射 $A^{(1-1)}$: $p(o^{(1-1)}\|s^{(1-1)})$	初始状态先验 $D^{(1-1)}$: $p(s^{(1-1)})$
1	直接注视	$\zeta_{1-1}=0.2$ $p(o^{(1-1)}\|s^{(1-1)})=\begin{bmatrix}0.691 & 0.309\\ 0.309 & 0.691\end{bmatrix}$	$p(s^{(1-1)})=\begin{bmatrix}0.5\\ 0.5\end{bmatrix}$
2	回避注视	$\zeta_{1-1}=0.2$ $p(o^{(1-1)}\|s^{(1-1)})=\begin{bmatrix}0.691 & 0.309\\ 0.309 & 0.691\end{bmatrix}$	$p(s^{(1-1)})=\begin{bmatrix}0.5\\ 0.5\end{bmatrix}$
3	直接注视	$\zeta_{1-1}=0\sim1.0$	$p(s^{(1-1)})=\begin{bmatrix}0.5\\ 0.5\end{bmatrix}$
4	直接注视	$\zeta_{1-1}=0.2$ $p(o^{(1-1)}\|s^{(1-1)})=\begin{bmatrix}0.691 & 0.309\\ 0.309 & 0.691\end{bmatrix}$	$a=0\sim1.0$ $p(s^{(1-1)})=\begin{bmatrix}a\\ 1-a\end{bmatrix}$

s：状态，o：观测，ζ_{1-1}：第1-1层似然映射的精度。上标表示变量所属的字段。

在条件1中，当智能体将高地位成员的直接注视作为观测 $o^{(1-1)}$ 时，我们模拟了动作选择概率。我们将观测模型的精度设为0.2，以反映推断状态与观察到的注视之间映射的不确定性。我们将先验设为 $p(s^{(1-1)})=\begin{bmatrix}0\\ .5\\ 0.5\end{bmatrix}$，假定智能体对高地位成员是否注意到自己以及是否对自己感兴趣作出相同的预测。

在条件2中，当智能体将高地位成员的回避注视作为观测 $o^{(1-1)}$ 时，模拟了动作选择概率。通过比较条件1和条件2的结果，预测了虚拟形象机器人将回避注视转换为直接注视对低地位成员陈述的影响。我们将

观测模型先验的精度设为与条件1相同的值。

在条件3下，通过改变观测模型的精度来模拟动作选择概率 $p\left(o^{(1-1)}|s^{(1-1)}\right)$。先验被设定为 $p\left(s^{(1-1)}\right) = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}^{\intercal}$，假设智能体会以相同概率预测高地位成员是否注意到自己并对自己感兴趣。研究预测了高地位成员将其注视行为改变为更加注意低地位成员所说内容并对此更感兴趣的行为所产生的影响。

在条件4下，通过改变先验 $p\left(s^{(1-1)}\right)$来模拟动作选择概率。观测模型的精度被设定为0.2，以考虑推断状态与观察到的注视之间映射的不确定性。先验编码了对高地位成员是否注意智能体并对其感兴趣的预测。研究预测了由虚拟形象机器人向高地位成员的注视行为添加先验动作 [17] 所产生的影响；该动作使得系统能够预测高地位成员针对低地位成员所说内容的注视移动。

表2列出了条件1和条件2下动作选择概率的仿真结果。在条件1下表达意见的概率高于条件2。因此，预测将高地位成员的注视行为从回避注视改为直接注视，会鼓励低地位成员发言。

图4显示了条件3下动作选择概率的仿真结果。观测模型的精度越高，表达意见的概率就越高。因此，预测将高地位成员的注视行为改为能够表达更多注意力和兴趣的方式，会鼓励低地位成员表达意见。

图5显示了条件4下动作选择概率的仿真结果。高地位成员将注意并对智能体表现出兴趣的预测程度越高，表达意见的概率就越高。基于这一结果，预测在高地位成员的注视行为中加入预备动作，将鼓励低地位成员发言。

基于这些发现，我们设计了虚拟形象机器人的注视行为。在本研究中，我们设计了远程参与者发言时的注视行为，这似乎表达了对本地参与者更多的注意。

表2. 条件1和条件2下动作选择概率的仿真结果。

s：状态，o：观测，ζ_{1-1}：第1-1层似然映射的精度。

条件	观测 $o^{(1-1)}$	似然映射 $A^{(1-1)}$: $p(o^{(1-1)}\|s^{(1-1)})$	初始状态先验 $D^{(1-1)}$: $p(s^{(1-1)})$	概率 of 表达意见
1	直接注视	$\zeta_{1-1} = 0.2$ $p(o^{(1-1)}\|s^{(1-1)}) = \begin{bmatrix} 0.691 & 0.309 \\ 0.309 & 0.691 \end{bmatrix}$	$p(s^{(1-1)}) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$	0.3729
2	回避注视	$\zeta_{1-1} = 0.2$ $p(o^{(1-1)}\|s^{(1-1)}) = \begin{bmatrix} 0.691 & 0.309 \\ 0.309 & 0.691 \end{bmatrix}$	$p(s^{(1-1)}) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$	0.0571

上标表示变量所属的字段。我们使用数学模型，模拟了在每种条件下，低地位成员看到高地位成员的注视行为后发言的概率。

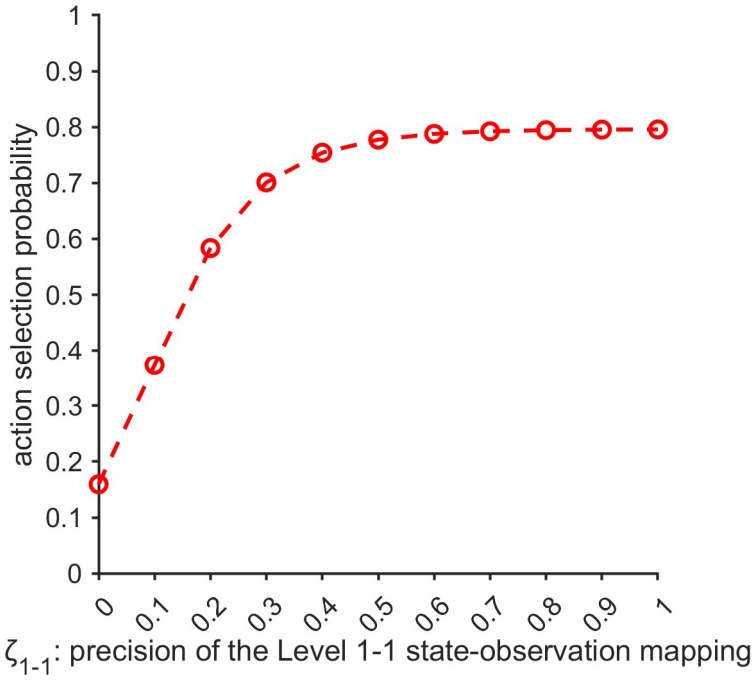


图4。第1-1层观测模型的精度对低地位成员表达意见的概率的影响。

精度会调节状态与观测之间映射的模糊程度。精度越高，表示映射越精确。第1-1层的观测模型是高地位成员的注意状态（注意/不关注低地位成员）与其注视行为之间的映射。

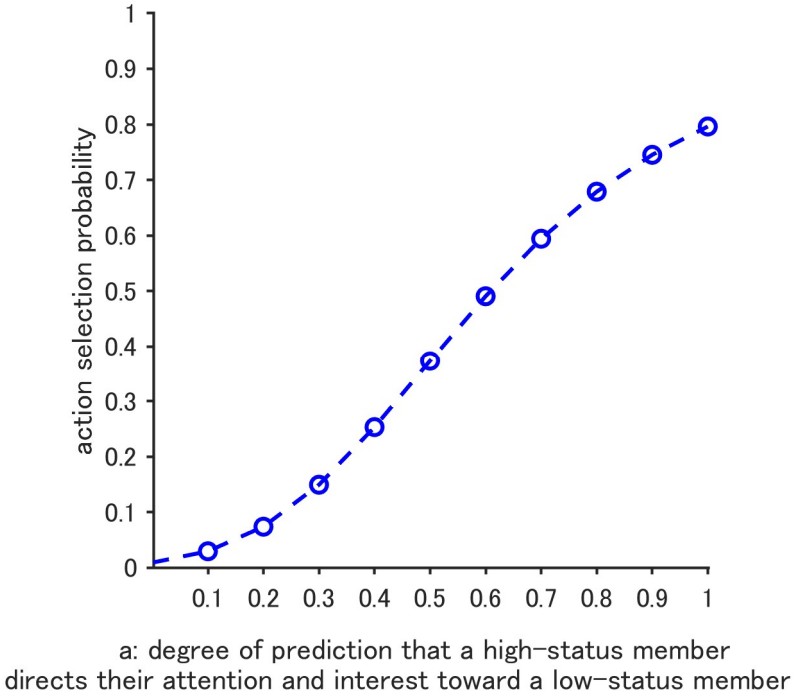


图5。第1-1层的先验分布（$p(s^{(1-1)}) = [a \quad 1-a]$）对低地位成员表达意见的概率的影响。

低地位成员在第1-1层推导出的状态是高地位成员的注意状态（注意/不关注低地位成员）。a 的值越高，表示低地位成员预测高地位成员将注意并对低地位成员产生兴趣的可能性越大。

虚拟形象机器人的注视行为设计

基于仿真结果，结合远程参与者的语音，对虚拟形象机器人的注视行为进行了设计。首先，说明注视行为的要求。其次，说明满足这些要求的注视行为设计方案。然后，说明实际设计的注视行为。

基于初步的模拟结果，我们决定让虚拟形象机器人自动进行直接注视，而不考虑远程参与者的注视行为。然而，如果注视行为表现得像是自动执行的，本地参与者可能无法感知远程参与者正在注意他们或对他们感兴趣。因此，第一项要求是，注视行为应当自然而非人为。我们还预测，在模拟中，表达出更多注意、兴趣和预备动作的注视行为会鼓励低地位成员发言。因此，第二项要求是，注视行为应当表达出更多注意和兴趣。我们使用了预备动作来满足这一要求。

接下来，我们将说明一种满足这两项要求的注视行为设计。本研究借鉴动画师的技巧和理念，设计自然的行为。动画师会通过强调和夸张必要的行为特征，并省略不必要的行为特征，来表现生动的行为 [17]。因此，我们的设计指南强调那些能够提示本地参与者注意的行为特征，并省略其余部分。这一指南很可能能够实现同时满足两项要求的注视行为。

最后，我们介绍设计的注视行为。表3列出了虚拟形象机器人的注视行为中包含的行为特征及其含义。远程参与者发言时，我们在注视行为中采用了这些行为特征。

表3. 融入虚拟形象机器人注视行为中的行为特征及其含义。

融入的行为特征 **虚拟形象机器人的注视行为**	**行为特征的含义**
头部摆动前后的眨眼	● 视点预备动作迅速的 [17][18]。 ● 模仿真实行为 [17]。 ● 强调头部摆动停止 [17]。
省略眨眼以外的动作 头部摆动前后	● 增强眨眼的印象头部摆动前后。 ● 增强直接注视的印象 [19]。

强调头部的慢入慢出 [17] 摆动		● 模仿真实行为。 ● 强调头部摆动停止 [17]。
眼球运动的实现 与眨眼和头部 摆动相关	眼球运动	● 真实行为的模仿。
伴随呼吸的头部摆动		● 真实行为的模仿。 ● 表达对象的 语音和眼部行为相同

实验：干预对虚拟形象机器人注视行为影响本地参与者陈述的效果

实验中使用的虚拟形象机器人的注视行为

参与者被分为实验组和对照组。远程参与者（研究人员）发言时，我们为两个组采用了不同的虚拟形象机器人注视行为。

对于实验组，使用了上一节所述的虚拟形象机器人。在此任务中，研究人员作为使用虚拟形象机器人的远程参与者，而两名参与者则是本地参与者。图6展示了实验中参与者与虚拟形象机器人的安排。虚拟形象机器人依次注视实验组中的每名参与者。当研究人员呼吸时，虚拟形象机器人的注视方向会发生变化。

在对照组中，虚拟形象机器人在研究人员进行语音讲述的整个过程中都注视着两名参与者的中间位置。为避免给人不自然的印象，我们设计并采用了一种眨眼模式：虚拟形象机器人在三秒后眨眼一次，又过三秒后眨眼两次。

注视行为的具体实现详见补充信息。

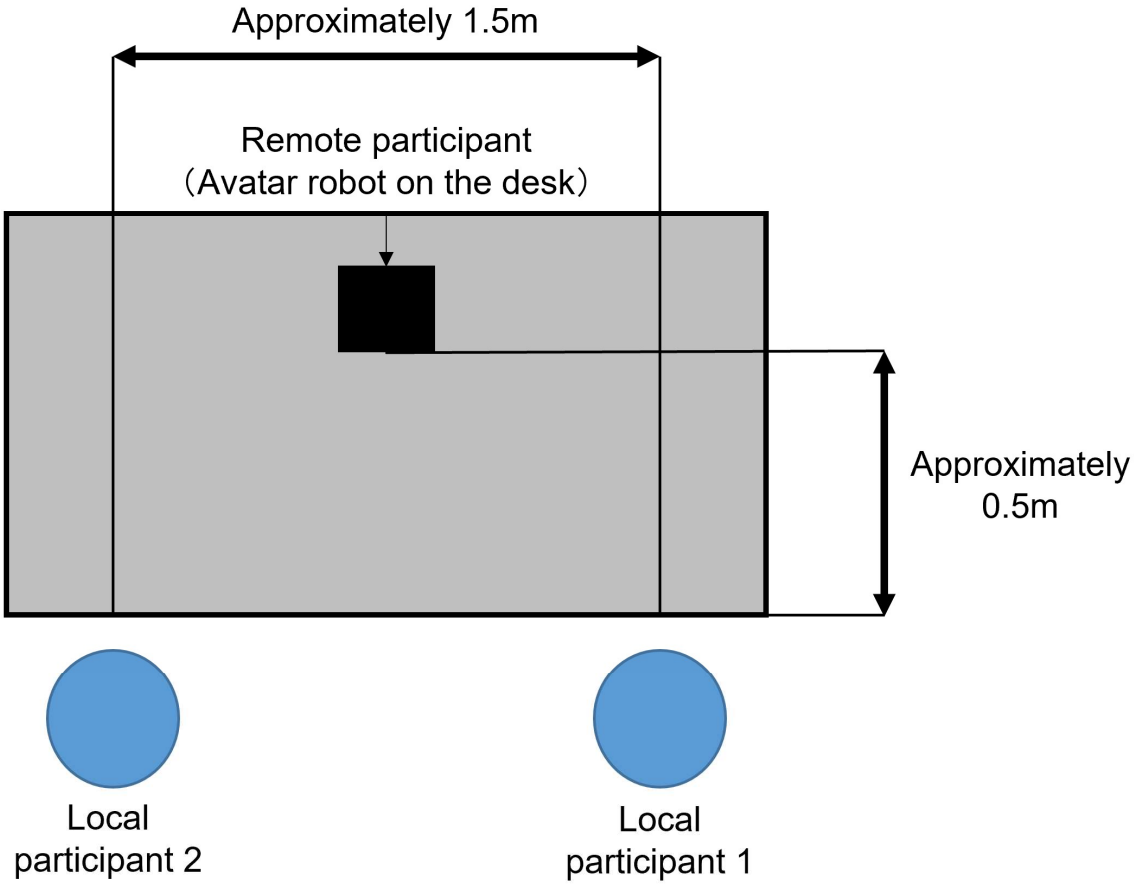


图6. 实验中参与者与虚拟形象机器人的布置。

参与者

参与者通过滚雪球抽样和自愿报名招募自东京大学的学生。这些学生没有视觉或听觉障碍，且母语为日语。实验组中有22名学生参与了本研究（男性16名，女性6名）；另有22名学生参与了本研究（男性16名，女性5名，1名未作答）。为控制熟人关系及性别组合的影响，我们将彼此不认识的同性参与者配对。一对彼此相识的女性参与者，以及一对在讨论前选择了相同候选人的女性和未作答参与者，均被排除在分析之外。

本研究经东京大学工学部研究伦理委员会批准（批准编号：KE22-59）。所有研究参与者均已提供知情同意。

设计

我们使用隐藏信息任务 [8] ，定量考察本地参与者的陈述是否得到了促进。在隐藏信息任务中，参与者会收到有关其选择的部分信息。只有彼此分享独特信息，参与者才能选出最佳选项。因此，隐藏信息任务用于研究群体信息 elaboration [13] 和学习中的团队绩效 [20]。

由一名研究人员和两名参与者组成的群体完成了一项隐藏信息任务。参与者阅读了三位学生会主席假设候选人的资料，然后决定哪位候选人最适合担任这一职位。资料没有包含有关候选人的完整信息，其中还包含仅提供给一名参与者的独特信息。我们通过评估分享的独特信息量，考察了每位参与者的陈述是否得到推广。

参与者进行了一项博弈，以再现隐藏信息任务讨论过程中老板与下属之间的关系。我们将其称为老板—下属博弈。研究人员以老板的身份参与博弈，而参与者则扮演下属。我们根据老板—下属博弈规则计算每位参与者的得分。随后，参与者被告知，除酬金外，他们还将根据自己的得分获得零食。实验结束后，所有参与者获得的零食数量均相同。老板—下属博弈规则规定，下属若不同意老板的意见，其陈述就存在被扣分的风险。这种风险与老板的主观判断有关。这种规则结构再现了一种情境：低地位成员由于面临人际风险而不愿表达自己的意见。

这些资料经过设计，使每位参与者分别支持不同的候选人：候选人 B 或候选人 C。研究人员作为主管支持候选人 B。因此，支持候选人 C 的参与者会根据老板—下属博弈规则承担其陈述带来的风险。这些资料是在修改 [8] 中所采用的方法后创建的。

在实际实验中，出现过参与者没有按照预期支持候选人的情况。在这种情况下，研究人员选择并支持候选人B或C，因此有些参与者赞同老板，有些则不赞同老板。我们将赞同老板的参与者称为参与者1，将不赞同老板的参与者称为参与者2。

步骤

完成知情同意书后，参与者填写了用于测试人格特质的问卷，然后开始进行隐藏信息任务。该隐藏信息任务包括个体阶段和讨论阶段。在个体阶段，每位参与者阅读资料，并指出自己当时支持哪些候选人。

然后，研究人员解释了在讨论阶段进行的老板—下属博弈。随后，研究人员移步到另一个房间，并连接到虚拟形象机器人。确认成功连接到虚拟形象机器人后，研究人员请参与者报告他们为个体阶段选择的候选人。报告顺序如下：收到推荐候选人 B 的资料的参与者先报告，随后是收到推荐候选人 C 的资料的参与者。

随后，研究人员解释了讨论阶段，并以老板的身份发表了意见（说明支持哪位候选人以及原因）。在研究人员进行解释时，实验组中的虚拟形象机器人依次注视每位参与者。在对照组中，当研究人员发表意见时，虚拟形象机器人继续注视两位参与者之间的中间位置。研究人员以老板的身份表达意见后，参与者就应提名哪位候选人担任学生会主席展开了自由讨论。在自由讨论期间，研究人员没有发言，虚拟形象机器人则继续观察两位参与者之间的中心位置。

自由讨论开始五分钟内达成共识后，参与者向研究人员报告了应被提名为学生会主席的候选人。研究人员收到报告后，隐藏信息任务即告完成。研究人员发放了一份问卷，要求参与者填写自由讨论期间群体分享的独特信息，以及对讨论阶段体验的主观评分。实验后说明并完成访谈后，参与者收到了零食作为额外奖励，随后离开。

测量指标

研究人员使用客观和主观指标评估虚拟形象机器人的注视行为对参与者的影响。

作为一项客观指标，测量了每位

参与者在自由讨论中分享的独特信息数量。参与者1（赞同将研究人员视为老板）分享的独特信息数量称为总和1。参与者2（不赞同将研究人员视为老板）分享的独特信息数量称为总和2。参与者1阅读的资料中包含六条支持老板意见的独特信息，以及三条不支持老板意见的信息。测量了支持老板意见的共享独特信息数量，该数量称为赞成1。这些独特信息表明老板支持的候选人值得选择，或表明老板不支持的候选人不值得选择。此外，还测量了不支持老板意见的共享独特信息数量，该数量称为不赞成1。赞成1和不赞成1回答的总数为总和1。参与者1根据所有独特信息，回答了有关其在自由讨论中与群体分享的独特信息的问题。参与者2阅读了九条不支持老板意见的独特信息。参与者2根据所有独特信息，回答了有关其在自由讨论中与群体分享的独特信息的问题。

此外，还测量了参与者犹豫分享的独特信息量。参与者1的回答数量称为犹豫1，参与者2的回答数量称为犹豫2。

使用了客观指标中未包含的主观指标，以衡量对参与者的影响。这些主观指标基于高阶成员的注视行为对低阶成员意见的影响模型。

测量了研究人员连接到虚拟形象机器人时，识别注视方向的难易程度。研究人员检查了参与者是否能够观察到模型第1-1层中高阶成员（即研究人员）的直接注视。

我们测量了研究人员连接到虚拟形象机器人时，参与者对关注和兴趣的感知。对关注和兴趣的感知反映了参与者是否推断自己正受到模型第1-1层中高阶成员的关注并引起其兴趣。对关注和兴趣的感知是一项指标，用于衡量参与者据以推断自己受到高阶成员关注并引起其兴趣的模型证据有多强（即高阶成员的临场感有多强）。

随后，我们测量了各群体的包容性。我们考察了第2层级的参与者是否根据自己对关注和兴趣的感知，推断自己已被纳入群体之中。

测量参与者在自由讨论中表达意见的难易程度，并检查参与者是否觉得在1-2级中选择表达自己的行动很容易。

参与者使用7级李克特量表对上述四个项目作答：识别注视方向的难易程度、对关注和兴趣的感知、参与群体的程度，以及表达意见的难易程度。

研究假设，结果会受到参与者人格特质的影响。因此，在隐藏信息任务之前，使用大五人格标记测量参与者的外向性和神经质[21]。

实验后说明结束后，研究人员询问参与者，他们是否认为在自由讨论中共享信息存在任何风险，以及研究人员发言时他们对虚拟形象机器人的印象如何。

此外，还录制了隐藏信息任务讨论阶段的视频，用于讨论实验结果。

假设

实验组采用的注视行为旨在向本地参与者表达注意力和兴趣。因此，我们提出以下假设：H1. 与对照组相比，实验组中参与者1（认同研究人员作为老板的参与者）和参与者2（不认同研究人员作为老板的参与者）识别注视方向的难度均有所降低。H2. 与对照组相比，实验组中参与者1和参与者2对关注和兴趣的感知均有所提升。根据高地位成员的注视行为对低地位成员陈述的影响模型，我们提出以下假设：H3. 与对照组相比，实验组中参与者2的包容性有所提升。H4. 与对照组相比，实验组中参与者2表达意见的难度有所降低。关于

参与者1在群体中的融入程度以及表达意见的难易程度，两个群体之间似乎没有显著差异。这是因为参与者1在两个群体中都认同老板，并对老板评价很高。

此外，我们提出以下假设：H5. 与对照组相比，实验组的指标 sum2 将增加。H6. 与对照组相比，实验组的犹豫指标 2 将减少。这是因为所设计的注视行为促进了

参与者2提出带有人际风险的陈述。相比之下，参与者1的陈述风险较低，因为其中包含大量支持老板意见的明确信息。在总和1和犹豫1方面，两组之间没有显著差异。

数据分析

我们分析了40名参与者的实验结果（实验组：16名男性和4名女性；对照组：16名男性和4名女性）。研究采用组间分析。当参与者的人格特质（外向性和神经质）与指标显著相关时，我们进行了协方差分析（ANCOVA），将人格特质作为协变量，并将参与者所属群体（实验组或对照组）作为因素。当参与者人格特质对指标的影响不显著时，我们采用双尾 t 检验比较实验组和对照组在客观评分和主观评分方面的差异。显著性阈值（显著性水平）设为0.05。

结果

图7显示了同意研究人员担任老板的参与者（参与者1）所分享的独特信息量。图8显示了不同意研究人员担任老板的参与者（参与者2）所分享的独特信息量。实验组的指标 sum2显著低于对照组（$t(18) = -3.104$, $p = 0.006$)。指标 sum2显示了参与者2所分享的独特信息量。该结果与H5相矛盾。此外，实验组的总和1值也显著低于对照组（$t(18) = -2.137$, $p = 0.047$)。指标总和1显示了参与者1所分享的独特信息量。

实验组与对照组之间没有显著差异

犹豫群体 2（$t(18) = 0.911$，$p = 0.828$）。因此，H6 未获支持。

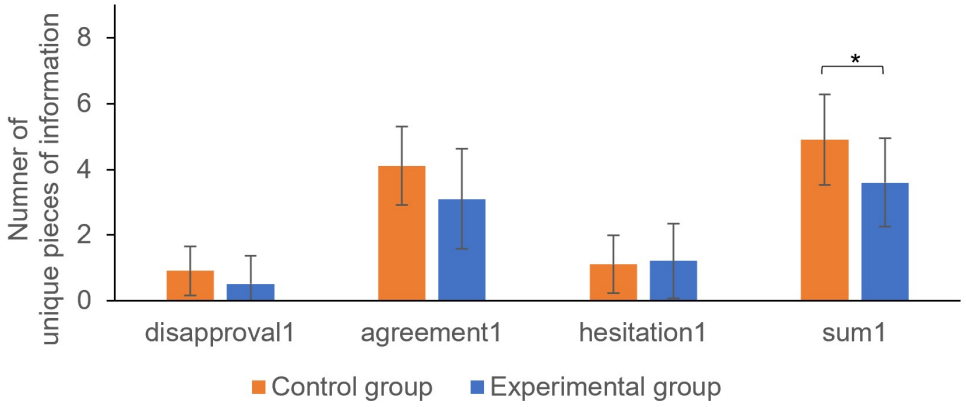


图7. 与高地位成员意见一致的参与者分享的独特信息条目数量。 * $p<0.05$，** $p<0.01$。误差线表示标准差。

不赞成1：分享的独特信息条目中，不支持高地位成员意见的条目数量。赞成1：分享的独特信息条目中，支持高地位成员意见的条目数量。犹豫1：参与者犹豫是否分享而未分享的独特信息条目数量。总和1：不赞成1和赞成1的总数。一名赞成高地位成员意见的参与者有六条支持该成员意见的独特信息条目，以及三条不支持该成员意见的独特信息条目。

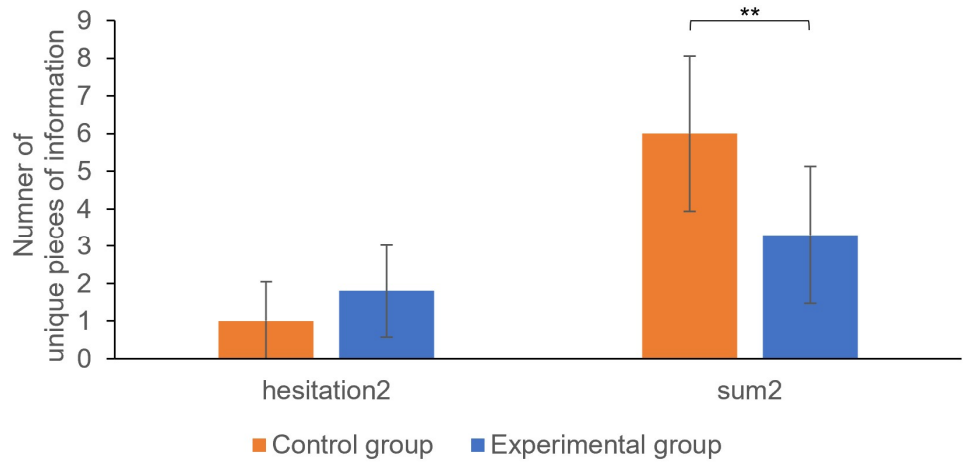


图8. 与高地位成员意见不一致的参与者分享的独特信息条目数量。 * $p<0.05$，** $p<0.01$。误差线表示标准差。

犹豫2：参与者在犹豫是否分享时未分享的独特信息条目数量。总和2：与高地位成员意见不一致的参与者所分享的独特信息条目数量。一位与高地位成员意见不一致的参与者拥有九条不支持该高地位成员意见的独特信息条目。

图9展示了参与者1的主观评分。图10展示了参与者2的主观评分。对于参与者1和参与者2，实验组识别注视方向的难易程度均显著高于对照组（参与者1识别注视方向的难易程度：$t(18) = 5.513$，$p < 0.001$；参与者2识别注视方向的难易程度：$t(18) = 2.994$，$p = 0.008$）。这些结果与H1一致。

对于参与者1和参与者2，实验组对关注和兴趣的感知均显著高于对照组（参与者1对关注和兴趣的感知：$t(18) = 2.832$，$p = 0.011$。参与者2对关注和兴趣的感知：$t(18) = 2.374$，$p = 0.029$）。因此，H2 得到支持。

然而，在群体内包容性或表达意见的难易程度方面，实验组和对照组之间没有显著差异（参与者1的包容性：$t(13.072) = -0.397$，$p = 0.698$；参与者1表达意见的难易程度：$t(18) = -1.698$，$p = 0.107$；参与者2的包容性：$t(18) = -0.616$，$p = 0.546$；参与者2表达意见的难易程度：$t(18) = -0.843$，$p = 0.410$）。因此，H3和H4未获支持。

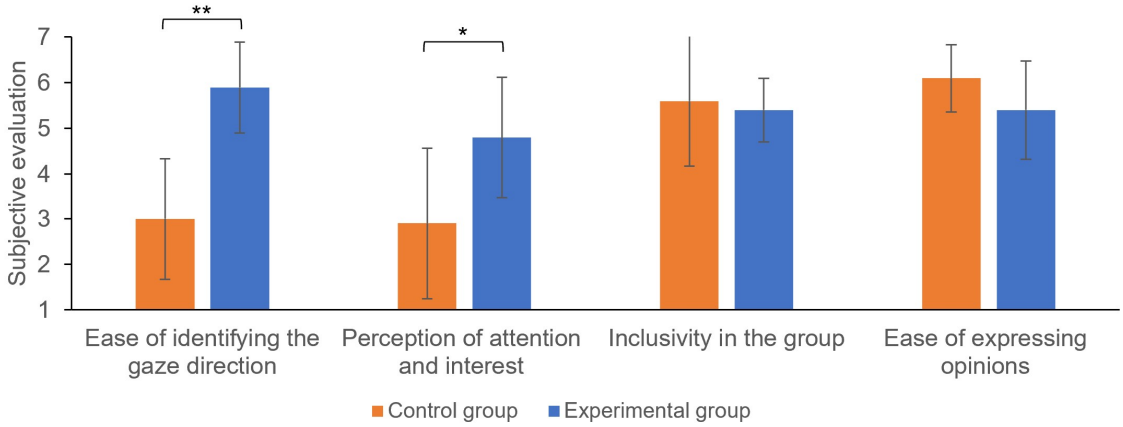


图 9. 与高地位成员意见一致的参与者的主观评价。 *p<0.05, ** p<0.01。误差条表示标准差。

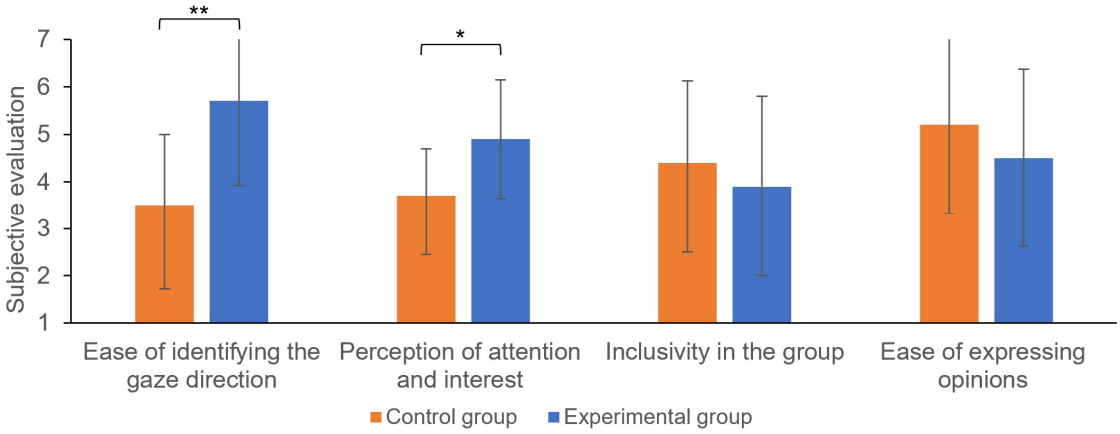


图10. 参与者对不同意高地位成员的参与者的主观评价。 * $p<0.05$, ** $p<0.01$。误差线表示标准差。

讨论

对于参与者1和参与者2而言，实验组对注意力和兴趣的感知均显著高于对照组（参与者1：认同研究人员是老板；参与者2：不认同研究人员是老板）。因此，正如假设所述，我们认为虚拟形象机器人的注视行为增强了远程参与者（作为老板的研究人员）的临场感。在事后汇报访谈中，我们还收到了参与者的反馈，例如“我觉得自己正在被注视”，以及“我以为研究人员正在和他注视的参与者交谈”。我们假设，注视行为通过让每位本地参与者推断老板正在关注他们并对他们感兴趣，增强了他们的模型证据，从而提升了远程老板的临场感。

然而，与 H5 相反，实验组的总和1和总和2显著小于对照组（总和1和总和2指标显示参与者1和参与者2分享的独特信息条目数量）。我们认为，这些结果是由于所构建模型对 Level2 的建模不正确所致。

在下属博弈中，如果参与者2表达意见，就可能被扣分。如果参与者2表达异议时，老板选择“拒绝意见”选项，参与者2的得分就会降低。因此，老板的存在可能使参与者2意识到自己正在受到.....

被老板监控，并强烈感受到表达意见的风险。换句话说，我们假设参与者2根据模型第2层级中老板的注意力和兴趣，推断出老板正在监控自己。

基于这项讨论，我们更新了高地位成员的注视行为对低地位成员陈述的影响模型，如图11所示。

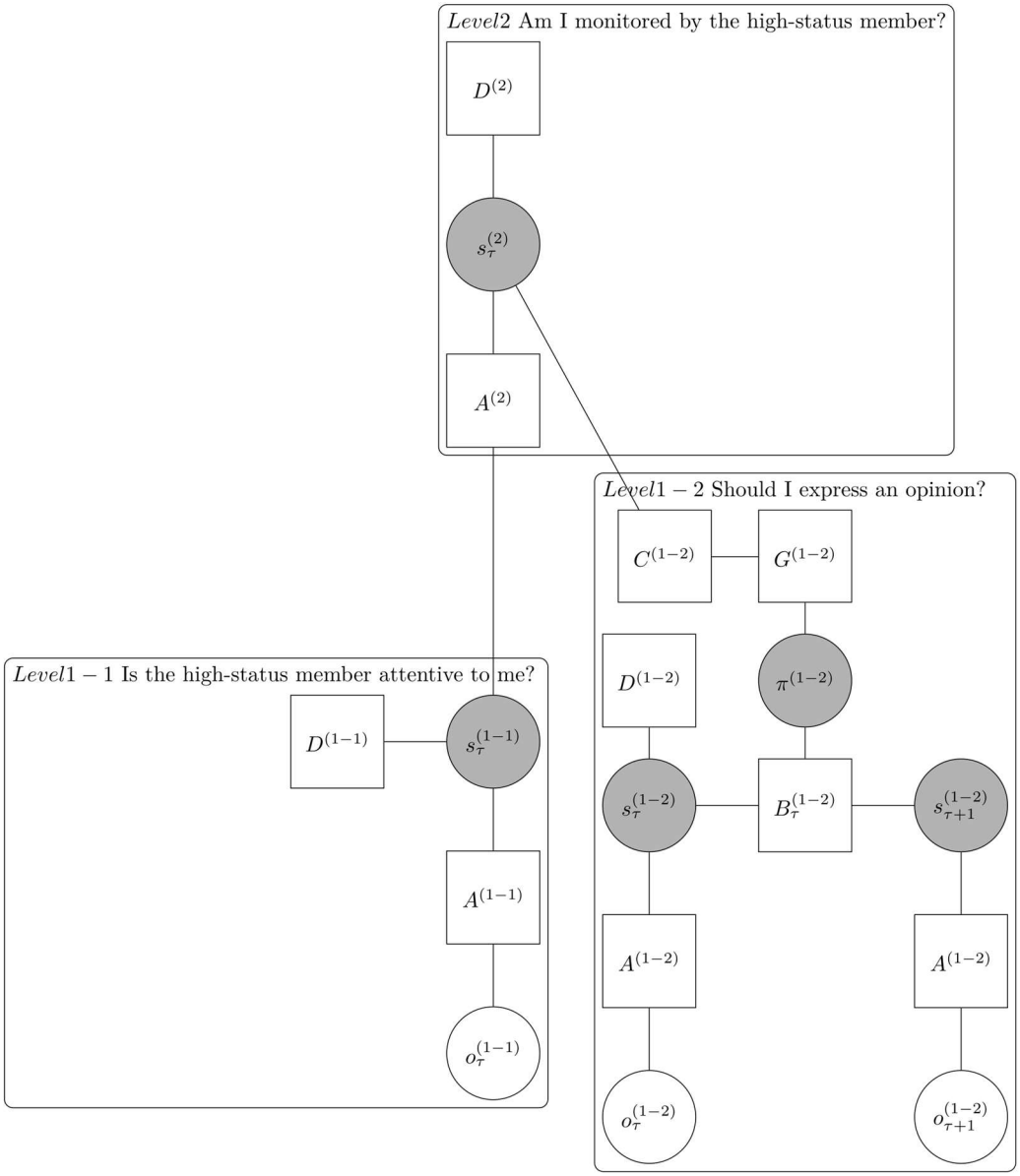


图11. 高地位成员的注视行为对低地位成员陈述的影响模型（修订版）。 第2层级中的推断状态和观测模型已修订。A：状态之间的似然映射以及

观测（观测模型），B：转移矩阵，C：先验偏好，D：初始状态先验，G：期望自由能，s：状态，o：观测，π：动作后验分布。上标用于指明变量所属的字段，下标表示时间步。在所有字段中，推断主体均为低地位成员。

首先，我们解释第1-1层。在这项实验中，正如假设的那样，虚拟形象机器人的注视行为提高了对关注和兴趣的感知。因此，第1-1层的观测模型定义为式（13），与修改前相同。

$$o^{(1-1)} = \{direct\ gaze \quad averted\ gaze\},$$
$$s^{(1-1)} = \{attentive \quad unconcerned\}$$
$$A_{1-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$p\left(o^{(1-1)} \middle| s^{(1-1)}\right) = \sigma(\zeta_{1-1} * \ln(A_{1-1} + \exp(-4))) \tag{13}$$

接下来，我们将描述第2层级。实验组和对照组在群体内包容性方面没有显著差异。不过，我们认为，对老板的关注和兴趣的感知增强了被老板监视的感觉。因此，我们将隐藏状态 $s^{(2)}$ 从 $s^{(2)} = \{included \quad excluded\}$ 改为 $s^{(2)} = \{monitored \quad not\ monitored\}$。我们将第2层级的观测模型重新表述为公式 (14)。

$$s^{(1-1)} = \{attentive \quad unconcerned\},$$
$$s^{(2)} = \{monitored \quad not\ monitored\}$$
$$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$p\left(s^{(1-1)} \middle| s^{(2)}\right) = \sigma(\zeta_2 * \ln(A_2 + \exp(-4))) \tag{14}$$

接下来描述 1-2级。1-2级的观测模型定义于式（15），与修改前相同。

$$o^{(1-2)} = \{agreement \quad disagreement\},$$
$$s^{(1-2)} = \{agreement \quad disagreement\}$$
$$p\left(o^{(1-2)} \middle| s^{(1-2)}\right) = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \tag{15}$$

由于情境 $s^{(2)}$ 发生了变化，我们修改了1-2级的先验偏好。在实验任务中，受到老板监视的低地位成员（参与者）会根据老板—下属

博弈规则感知被扣分的风险。因此，老板—下属博弈规则决定了不同意老板所带来的风险。因此，我们将 $s^{(2)}$ 从 $s^{(2)} = \{included \quad excluded\}$改为 $s^{(2)} = \{monitored \quad not\ monitored\}$，并且仅在 $s^{(2)} = monitored$时对 $o^{(1-2)} = disagreement$ 设定负偏好。我们将1-2级的先验偏好重新表述为式（16）。

$$o^{(1-2)} = \{agreement \quad disagreement\},$$
$$s^{(2)} = \{monitored \quad not\ monitored\}$$
$$p\left(o^{(1-2)}\middle|s^{(2)}\right) = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} \tag{16}$$

使用更新后的模型进行了模拟。我们改变了$o^{(1-1)}$ 在各条件下的观测。表9列出了模拟条件以及动作选择概率的结果。各条件下未发生变化的模型变量均设置为与修改前模型中的变量相同，具体内容见补充信息。

表9. 使用修改后的模型进行仿真的条件，以及各条件下动作选择概率的仿真结果。

条件	观测 $o^{(1-1)}$	似然映射 $A^{(1-1)}$: $p\left(o^{(1-1)}\middle\|s^{(1-1)}\right)$	初始状态先验 $D^{(1-1)}$: $p(s^{(1-1)})$	概率 of 表达意见
1	直接注视	$\zeta_{1-1} = 0.2$ $p\left(o^{(1-1)}\middle\|s^{(1-1)}\right) = \begin{bmatrix} 0.691 & 0.309 \\ 0.309 & 0.691 \end{bmatrix}$	$p\left(s^{(1-1)}\right) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$	0.0571
2	回避注视	$\zeta_{1-1} = 0.2$ $p\left(o^{(1-1)}\middle\|s^{(1-1)}\right) = \begin{bmatrix} 0.691 & 0.309 \\ 0.309 & 0.691 \end{bmatrix}$	$p\left(s^{(1-1)}\right) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$	0.3729

s：状态，o：观测，ζ_{1-1}：第1-1层似然映射的精度。上标表示变量所属的字段。

表9中的仿真结果表明，在

条件1低于条件2，因此，向低地位成员直接注视的高地位成员（或虚拟形象机器人）会 withholding 他们的陈述。更新后的模型模拟支持了实验结果，即实验组中的总和2低于对照组中的总和2。

我们认为，参与者2不作陈述导致实验组中的总和1减少。当参与者2话不多时，参与者1也无需就物体作过多陈述。此外，总和2与总和1之间存在显著正相关（$r = 0.587$, $N = 20$, $p = 0.007$）。

虚拟形象机器人的注视行为仅在自由讨论开始前进行，在自由讨论中，参与者表达了自己的意见。尽管如此，注视行为显著增强了远程参与者的临场感，并影响了本地参与者在自由讨论中的陈述。

结论

在本研究中，我们将临场感表述为推断他人心理状态的模型证据。随后，我们提出，注视行为应能增强临场感（即模型证据），从而推断远程参与者正在关注本地参与者。基于该模型的模拟预测，注视行为会促进本地参与者发表陈述。我们采用会议任务，重现了立场不同的群体之间的讨论，以探究所设计的注视行为是否能促进本地参与者表达意见。实验结果表明，虚拟形象机器人的注视行为显著增强了远程参与者的临场感，从而抑制了本地参与者在实验任务情境中表达意见。

所提出的虚拟形象机器人的注视行为增强了远程参与者的临场感。我们认为，在远程参与者的临场感会产生积极影响的情境中，注视行为有助于沟通。例如，当下属想要展现自己的技能或为老板作出贡献时，老板在场会鼓励下属发表陈述。在这些情况下，将老板的行为转化为本研究开发的虚拟形象机器人的注视行为，可能会促进下属发表陈述。

在此情境下，发言时表现出较少注意的行为会促进下属陈述，这一发现也可用于通过虚拟形象机器人进行干预。虚拟形象机器人

的行为可以在此情境下发生变化。在老板的临场感带来压力的情境中，虚拟形象机器人可以表现出较少注意的行为；而在老板的临场感鼓励下属的情境中，则可以表现出本研究所提出的注视行为。因此，了解远程参与者在不同情境下恰当的临场感非常重要。

实验结果表明，远程参与者的临场感抑制了本地参与者的陈述。该结果与我们的假设不一致。然而，这表明模型的临场感具有足够的影响力，能够限制参与者的陈述。此外，先前的研究[1][2] 采取了增强临场感的方法。然而，正如本研究所显示的，临场感可能会对沟通产生负面影响。因此，我们认为，使用数学模型研究临场感的影响及其控制方法是有价值的。

本研究存在一些局限性。首先，实验仅以日本参与者为对象进行。虚拟形象机器人的注视行为在不同文化中可能产生不同的影响。其次，本实验未对参与者关于机器人特征的先验知识进行标准化。我们假设，当参与者相信机器人反映了远程参与者的行为时，与不相信时相比，他们对该行为的感知会有所不同。

.

致谢

我们谨向东京大学工学部机械工程系的村上保教授表示诚挚的感谢，感谢他在研讨会期间提出的宝贵意见。

我们衷心感谢铃木隆一、安田真理、西田直树、新城真以及索尼集团的其他员工为我们的研究活动提供必要的设备，就用于实现虚拟形象机器人行为的代码提供咨询，并通过会议和虚拟形象机器人体验提出宝贵建议。

我们衷心感谢首席技术专家及川和弘，就设施、设备和实验环境向我们提供建议，并为我们准备研究环境。

我们谨向秘书野崎由里致以诚挚谢意，感谢她提供帮助

协助处理实验参与者的报酬相关文书工作。

我们还要衷心感谢所有参与本研究预实验和正式实验的人员，以及帮助我们招募参与者的人员。

参考文献

1. Yankelovich N, Simpson N, Kaplan J, Provino J. Porta-person：互联会议室的远程呈现。ACM；2007年。https://doi.org/10.1145/1240866.1241080 2. van Dijk B, Zwiers J, op den Akker R, Kulyk O, Hondorp H, Hofs D, 等。通过传达方向性注视线索来支持混合会议中的远程参与。载于：迈向自主、自适应和上下文感知的多模态界面。理论与实践问题。柏林，海德堡：Springer Berlin Heidelberg, 第412-28页。https://doi.org/10.1007/978-3-642-18184-9_36 3. Ogasawara KJ, Usuda T, Wu X, Hasegawa T, Hirako M, Fujita S, 等。视觉-嗅觉整合以丰富数字体验。国际情感科学与工程研讨会。2022年；ISASE2022(0):1–4。https://doi.org/10.5057/isase.2022-C000033 4. Kawaguchi I, Morita T, Kuzuoka H, Suzuki Y。具有类人头部的远程呈现机器人所呈现的注视对远程通信的影响。人机交互学会会刊。2018年；20(4):457-68。5. Heylen DKJ。头部动作、注视与会话结构原则。国际人形机器人期刊。2006年；3(11/3):241-67。https://doi.org/10.1142/S0219843606000746 6. Hall JA, Horgan TG, Murphy NA。非言语沟通。心理学年度评论。2019年；70(1):271-94。https://doi.org/10.1146/annurev-psych-010418- 103145 7. Smith R, Friston KJ, Whyte CJ。主动推断及其在实证数据中应用的分步教程。数学心理学杂志。2022年；107:102632。https://doi.org/10.1016/j.jmp.2021.102632 8. Stasser G, Titus W。群体决策中未共享信息的汇集：讨论期间的偏差信息抽样。人格与社会心理学杂志。1985年；48(6):1467-78。https://doi.org/10.1037/0022-3514.48.6.1467 9. Mitchell R, Nicholas S, Boyle B。对认知多样性的开放性与群体的作用

知识创造过程。《小组研究》。2009年；40(5):535-54。
https://doi.org/10.1177/1046496409338302 10. Schulz-Hardt S, Brodbeck FC, Mojzisch A, Kerschreiter R, Frey D。隐藏信息情境中的群体决策：异议有助于提升决策质量。《人格与社会心理学杂志》。2006年；91(6):1080-93。
https://doi.org/10.1037/0022-3514.91.6.1080 11. Garner JT, Iba DL。眼神接触和吸引力评分相对于排斥与异议的变化。《小组研究》。2015年；46(1):3-26。
https://doi.org/10.1177/1046496414550880 12. Edmondson A。工作团队中的心理安全感与学习行为。《行政科学季刊》。1999年；44(2):350-83。
https://doi.org/10.2307/2666999 13. Shim S, Livingston RW, Phillips KW, Lam SSK。领导者目光对成员影响力差异的影响：对多元化群体中过程与绩效的启示。《管理学会期刊》。2021年；64(6):1873-900。
https://doi.org/10.5465/amj.2017.1507 14. Frischen A, Bayliss AP, Tipper SP。注视线索对注意的引导：视觉注意、社会认知与个体差异。《心理学公报》。2007年；133(4):694-724。https://doi.org/10.1037/0033-2909.133.4.694 15. Hirak R, Peng AC, Carmeli A, Schaubroeck JM。领导者包容性与工作单元绩效的关联：心理安全感和从失败中学习的重要性。《领导力季刊》。2012年；23(1):107-17。https://doi.org/10.1016/j.leaqua.2011.11.009 16. Parr T, Friston KJ。推断的解剖学：生成模型与脑结构。《计算神经科学前沿》。2018年；12:90-。https://doi.org/10.3389/fncom.2018.00090 17. Richard W。Yoko G 译。《动画师生存手册：扩充版》。Graphic-sha出版社；2011年。18. Takafumi M。动画作为“用时间作画”：庵野秀明执导的《福音战士新剧场版：破》（2009）中的艺术表达。《人文论究》。2012年；61(4):99-125。19. Omori Y, Miyata Y。基于眨眼频率对印象的估计。《社会行为与人格》。2001年；29(2):159-67。
https://doi.org/10.2224/sbp.2001.29.2.159

20. Tost LP, Gino F, Larrick RP. 权力使他人沉默：领导者权力对团队绩效的负面影响。《管理学会期刊》。2013年；56(5)：1465-86。
https://doi.org/10.5465/amj.2011.0180 21. Tsutomu N, Iori T, Takafumi W, Ryuichi K, Ai N, Hiroyuki N. 日本大五人格量表短版的开发及其信度和效度检验。《日本心理学杂志》。2012年；83(2)：91-9。
https://doi.org/10.4992/jjpsy.83.91

补充信息 1

2 **S1 文本。虚拟形象机器人的注视行为具体实现** .

收集了伴随注视行为的头部旋转、注视和眨眼变化数据。根据具体情境对数据进行了分析。S2表 4 展示了已实施注视行为的流程。5

此处介绍行为6中使用的头部旋转、注视和眨眼数据。行为6中使用的头部旋转、注视和眨眼图表见S3–S5图。

头部旋转数据由三个角度组成：偏航（摇头）、俯仰（点头）和横滚（倾斜）。三轴执行器根据这三个角度旋转。在行为6中，俯仰角为0°，与结合注视时和本地参与者进行眼神接触的角度相对应。横滚角保持在0°不变，表示头部没有倾斜。行为开始后0.52秒，偏航角从-20°（与本地参与者2进行眼神接触时的角度）变为10°（与本地参与者1进行眼神接触时的角度）。由于眼睛的设计，与本地参与者1进行眼神接触时的角度绝对值小于与本地参与者2进行眼神接触时的角度。

使用反正切函数生成了 0.52 s 内偏航变化的数据，如公式（S1）所示。18

$$\omega = 10^{\frac{2t-0.5}{0.25}}$$

19

$$yaw = \begin{cases} 5, \ if\ \omega = 1, \\ tan^{-1}\dfrac{2\zeta\omega}{1-\omega^2} - 20, if\ \omega \neq 1 \end{cases} \quad (S1)$$

t 表示从行为开始起经过的时间，ζ 是用于调整慢入慢出侧重程度的参数。公式（S1）第三行的第一项，是角频率为 ω ($G(s) = \frac{1}{s^2+2\zeta s+1}$) 的二阶延迟元件相位角的反号。利用公式（S1），可以在保持偏航角变化所需时间不变的同时，调整慢入慢出的侧重程度。25

注视数据表示为瞳孔在 xy- 26 平面上的坐标位置，其中 0≤x≤100 和 0≤y≤100。瞳孔图像会移动到眼部显示器上由 27 x–y 坐标定义的位置。x坐标（水平坐标）

沿视线转移方向线性移动，然后在短于偏航角变化的时间（0.31 秒）内稍微向相反方向返回。29 这表明眼睛的移动早于头部转动。垂直坐标（y 坐标）会在眨眼时向下移动。31 这些眼球运动模拟了通过分析从虚拟形象机器人获取的眼动数据日志所发现的行为特征。33 眨眼数据是一个介于 0 到 100 之间的整数，表示眼睛睁开的程度。34 数值为 0 时，虚拟形象机器人的眼睛处于闭合状态；数值为 100 时，眼睛完全睁开。35 眨眼发生在偏航变化开始时，以及偏航变化结束前一刻。36 具体而言，眨眼分别发生在头部摆动之前和之后。37 其他行为的数据通过应用行为6的数据生成。38 如表 S5 所示，虚拟形象机器人会在远程参与者语音开始时、呼吸期间以及语音结束时改变行为。40 我们将语音开始编码为远程参与者输入音频的 16 位 PCM 值超过某一阈值的时刻。41 屏息时机编码为行为 2–6 期间，远程参与者输入音频的 16 位 PCM 值低于某一阈值的时刻。43 远程参与者语音结束的定义为参与者按下按钮的时刻。45

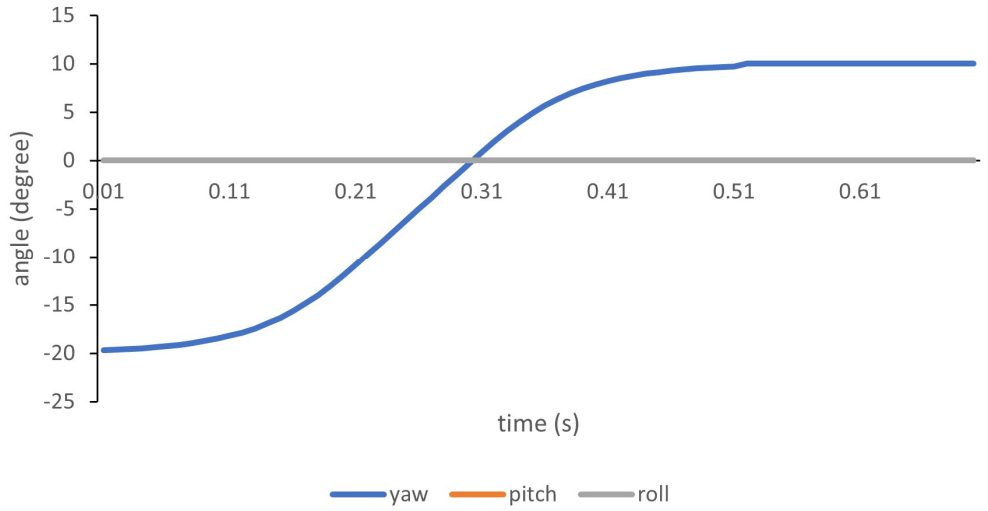
46

S2表。虚拟形象机器人的注视行为流程。47

注视行为发生的时机	**行为编号**	**注视行为**
远程参与者发言前。 （在受控群体中，虚拟形象机器人在研究人员 发言时执行行为 1。）	1	继续在两位 本地参与者。
当一位远程参与者开始发言时。	2	将注视转向本地 参与者1。
行为2或行为6之后，直到一名 远程参与者吸气。	3	持续注视本地 参与者1。
当远程参与者吸气时 在行为3期间。	4	将其注视方向指向本地 参与者 2。
在行为4之后，直到一位远程参与者 深吸了一口气。	5	持续注视局部地方 参与者 2。

当远程参与者在行为 5 期间呼吸时。	6	将视线转向本地参与者1。
在远程参与者的语音结束时。如果虚拟形象机器人处于行为3，则启动行为7；如果虚拟形象机器人处于行为5，则启动行为8。	7,8	将其注视方向在两位本地参与者之间切换。

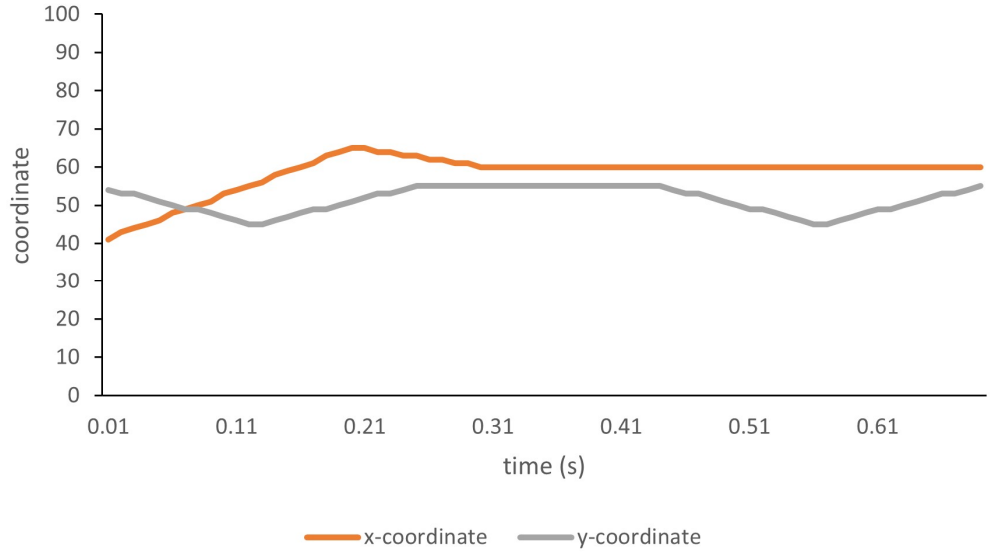
48



49

50 **S3 图。行为6中虚拟形象机器人的头部旋转。**将这些数据作为输入，我们控制虚拟形象机器人的头部旋转。51

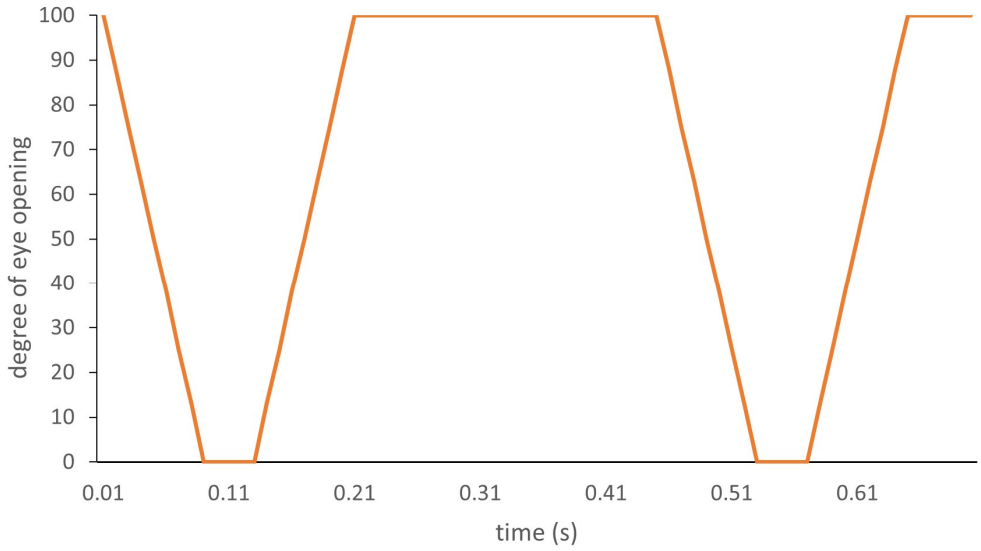
52



53

S4 图。行为6中虚拟形象机器人的注视。注视数据为瞳孔位置 54

表示为 xy平面上的坐标，其中 0≤x≤100， 0≤y≤100。使两只眼睛的瞳孔位置保持一致。



S5 图. 行为6中虚拟形象机器人的眨眼。眨眼数据为 0 到 100 的整数，表示眼睛睁开程度；当数值为 0 时，虚拟形象机器人的眼睛闭合，当数值为 100 时，眼睛则完全睁开。