arXiv:2302.01854v1 [cs.HC] 3 Feb 2023

比较人机交互中依从性的心理测量与行为预测因素 *

Nikolos Gurney[1 [0000−0003−3479−2037]]、David V. Pynadath[1, 2 [0000−0003−2452−4733]]和 Ning Wang[1,2]

[1] 创意技术研究所 gurney,pynadath,nwang@ict.usc.edu[2] 维特比工程学院，计算机科学系 南加州大学，美国加利福尼亚州洛杉矶

摘要。 人机团队的优化取决于人工智能能否针对不同的人类队友调整交互方式。自适应人工智能研究中的一个常见假设是，人们信任倾向上的细微差异会显著影响他们遵从人工智能建议的可能性。信任倾向通常通过在交互前施测的自我报告量表来测量。我们将一种广受欢迎的此类测量工具与依从性的行为预测因素进行基准比较。我们发现，在取自三个既往研究项目的数据集中，该量表对依从性的预测效果不如行为测量。这表明，初始行为上的个体差异通常比自我报告的信任态度差异更具预测力。这一结果还表明，易于获取的行为测量有望在无需使用（通常成本高昂的）调查工具的情况下，为人工智能提供更准确的模型。

关键词： 人机交互 · 人机交互 · 依从性 · 信任 · 干预 · 决策。

1 引言

准确建模人们对人工智能的态度（例如信任）的能力，对于设计成功的人机交互至关重要。Leeand See [21] 在其关于对自动化的信任的开创性著作中指出，人们在信任倾向上的细微差异，可能导致他们与自主系统产生显著不同的交互。例如，信任倾向较高的人可能会冒险依从

* 本研究的部分工作由美国国防高级研究计划局（DARPA）根据合同编号 W911NF2010011 资助。本文内容未必反映美国政府或美国国防高级研究计划局的立场或策略，也不应据此推断其给予了官方认可。

系统的建议，而信任倾向没那么强的人则不会。其假设是，如果这个自主系统足够可靠，那么依从的队友很可能会获得积极结果；又由于假定积极结果能够预测信任的提升，这种经历反过来应会带来更多依从性。因此，依从这一行为会启动一个反馈循环，让信任滋生信任。然而，信任倾向较低的人不太可能依从，随后也就不太可能进入这一积极反馈循环。

具备建模能力并能将个人的先天倾向（无论是否与信任有关）纳入决策的自适应人工智能，或许能够避免人们对其持有的校准失当态度所带来的负面影响。以不服从的队友为例，这种人工智能可以在提出建议时附上更深入的解释和理由，旨在克服此人信任倾向不足的问题。对建议进行这样的个性化处理，需要持续测量并建模信任—服从关系。尽管 Lee 和 See 提出了人类与自动化系统团队中信任影响的模型，但并未提供测量方法。

研究人员投入了大量资源，研究 Lee 和 See 所假设的信任倾向与对自动化的信任之间的关系，包括如何测量并据此进行调整 [17]。一种常见做法是使用通用量表测量个人的信任倾向，然后将该测量结果与其在和自动化系统交互时的人类行为相关联。这类量表中一个广受欢迎的例子是包含 12 个条目的信任倾向量表（DTI） [23]。该测量旨在捕捉个人总体上依赖他人的可能性，并假设其会影响个人对他人以及其他受其信任影响的系统所持有的信念和意图。有证据表明，该测量在与虚拟机器人和实体机器人交互时具有（例如 [33,2]）预测价值，也有证据表明其不具有（例如 [44,34]）预测价值。

如今，了解信任倾向等个体特征如何影响依从性，比以往任何时候都更加重要。研究人员、技术专家和资助机构正越来越多地致力于开发能够与人类进行丰富交互的自动化系统（例如 [4,13,37,38]）。尽管付出了这些努力，人工智能智能体仍会因其运行环境的复杂性和固有不确定性而不断出错 [4]。一种减轻这些错误所造成损害、进而促成更恰当校准的信任的方法，是让智能体以符合人类队友不同信息需求和偏好的方式传达其推理。沟通具有充分的透明度，能让人们知道何时应依从人工智能的建议、何时不应依从，即形成校准得更恰当的信任；而这正是人机交互中良好依从行为的重要前提 [21,8,44,33,2]。

另一方面，如果智能体对人们不同的信任倾向不敏感，就会对所有人采用相同的沟通策略。鉴于个体差异无处不在，这种一刀切的策略必然会导致

对智能体 [30]的不当误用或弃用。因此，尽可能早地预测信任——或许甚至将其视为一种特质，正如 DTI 所致力于做到的那样——就显得尤为重要。例如，如果机器人能在交互开始前就建立起对新的人类队友信任倾向的模型，就能抢先个性化调整沟通方式，以最大限度地提升透明度、信任校准和适当的依从性。由此可见，我们迫切需要依从性预测基准：它们不仅要准确，还应能让智能体尽可能早地获取，同时不牺牲预测价值。

我们将 DTI 与四种简单的行为测量进行比较：一个人是否采纳机器人的第一条建议、机器人第一次出错后立即给出的建议、机器人在第一次出错前提出的所有建议，以及机器人在早期任务过程中给出的建议。我们使用三个独立实验的数据（这些实验旨在研究人机交互），比较这些测量预测客观（行为）结果的能力。在全部三个数据集中，与行为测量相比，DTI 对人机团队绩效方差的解释程度更低（因此预测能力也更弱）；而且，与 [23],所假设的效应相反，它无法预测初始依从性或总体服从。在三个数据集中的两个数据集中，DTI 甚至无法解释模型中显著比例的方差，这意味着按当前建模方式，它不具备预测价值。在第三个数据集中，DTI 与依从性呈负相关。也就是说，尽管假设较高的 DTI 得分能够预测更高的依从性，但它实际预测的结果恰恰相反。鉴于在许多人类与人工智能交互领域，开展 DTI 这类问卷调查并不切实际（有时甚至不可能），令人欣慰的是，这些早期行为测量（智能体通常可以轻易观察到）能够提供如此强的预测能力。事实上，假设关于依从性与信任之间关系的假设是准确的，那么与基于自我报告的 DTI 这类测量相比，这些行为测量似乎更能准确反映一个人真实的信任倾向。

我们相信，这一相对简单的发现将为一系列多样化的人机交互领域中的模型提供有价值的输入。诚然，现有许多用于预测依从性的主观信任测量指标，但实施这些测量通常负担较重，而且其预测效果并不总如预期。同样，也有许多用于预测依从性的信任计算模型，但它们大多针对特定领域，且几乎总是难以实施。相比之下，我们的行为方法易于评估、实施简便，更重要的是，还能预测依从性。此外，它有望帮助开发更稳健的依从性预测指标，识别何时可能需要采取依从性干预，并为其他类似基准提供原型。我们认为，识别并开发易于评估的测量方法（例如本文所介绍的信任行为测量）是开发稳健、自适应且能够实施个性化干预的人工智能的关键步骤。

2 相关工作

“信任”通常被用作一个总括性术语，涵盖影响依从性的各种人为因素（有关这些因素的分类荟萃分析，请参见 [16]）。在人工智能研究中，对其中一些特定信任因素进行测量和建模是一种常见做法。除前述DTI之外，研究人员还开发或改编了许多其他测量工具，以便理解这些因素，但通常彼此孤立地进行研究。人工智能研究人员采用这些测量工具，通常反映了每个目标因素的细微差异。研究人员可以使用这些广义测量指标，为每个人建立假设基线；而对于其他目的，则可以采用更具体的测量指标，正如研究人员可以使用DTI，为某个人的信任倾向提供基线一样。全面考察所有这些因素和测量工具超出了本研究的范围，因此我们将重点放在与个人先验地信任自主系统的倾向相关的因素和测量工具上。

DTI旨在识别人类队友固有的信任水平，而信任倾向量表旨在确定他们是否倾向于信任 [6]。与其实施相关的假设类似：与信任倾向较低的人相比，信任倾向较高的人更有可能遵从自动化系统的建议。研究人员已将这一测量作为控制个体对机器人社会工程响应中方差的一种手段 [5,3] ，或用于理解注意力控制与信任之间的关系 [42]。

交互的情境可能会改变一个人对自动化系统的反应。考虑到这一点，研究人员针对众多用例开发并采用了信任测量。在一个抽象层次上，有些量表旨在评估人们对特定类型自动化的信任，例如人机交互信任量表，顾名思义，该量表用于评估一个人对机器人 [49] 的信任程度；还有人机信任量表，该量表旨在评估一个人对机器人信任的变化 [35]。研究人员进一步针对特定用例改进了此类量表、DTI及其他量表。例如，社会服务机器人交互信任量表的设计重点，就是服务情境下的人机交互 [11]。

最后，还有一些计算方法可用于追踪在与机器人进行长时间交互时信任的演变。这些方法，例如 [48] 中提出的概率模型或 [32]，中介绍的马尔可夫模型，能够让机器人动态调整其对人们赋予自身之信任的表征。通常，这类方法会以类似信任量表（如人机信任量表）的结果作为初始依据，但无需重复测量。

3 信任倾向的测量

McKnight 等人 [23] 开发了 DTI（见附录 A 的图 1），作为一个更大构念的一部分，专门用于电子商务，以预测

与供应商初次交互的结果。它本身包含四个子测量维度：仁慈、诚信、能力和信任倾向（每个维度分别包含三个条目）。通常，评估会要求人们使用反应量表表明自己对每条陈述的认同程度。随后将这些反应取平均值，并与感兴趣的结果测量相关联，例如依从性。其基本假设是，如果一个人在 DTI 上得分较高，那么他们就更可能信任某个供应商（向其分享私人信息、向其购买商品、听取其建议等），从而依从其请求。

将 DTI 等信任测量工具应用于人机交互或类似情境时，隐含的假设是，该工具仍然可靠，并且在开发过程中观察到的构念效度在新领域中依然成立。截至本文撰写时，介绍该测量工具的论文已被引用逾 6000 次，其中包括大量在人机交互中的应用（例如 [44,45,33,46,22,34,41,2,3,31]）。不过，未发现任何验证该量表适用于人机交互情境的出版物。

此外，值得注意的是，原作者并未根据实际的信任行为对 DTI 进行验证。正如他们所指出的，在电子商务中常见的信任相关行为包括分享个人信息、进行购买，或根据网站提供的信息采取行动。由于成本原因，作者未采用上述任何行为来验证其测量工具。相反，他们要求研究参与者指出自己做出各种信任行为的可能性；作者援引先前研究为此做法辩护，称对这类测量的回答与实际行为并无显著差异（例如 [43]）。总而言之，该量表旨在研究电子商务情境中的购买意向，而非人机交互中的依从潜力。

这两点——该量表是在不同情境下开发的，且旨在预测意图而非行为——表明它可能不适合用于构建预测人机交互中依从性的模型。还有其他可用的测量工具，例如[44] 使用的信任倾向量表，以及近期开发的用于预测对自动化的信任的测量工具 [18,24]„ 但它们的应用有限，而且同样并非专为人机交互情境设计。此外，还有一些针对特定态度开发的测量工具，例如对机器人负面态度量表 [28]„ 但同样并非专为预测依从性而设计。这些都表明，我们需要采用其他方法，为构建能够预测人机交互中依从行为的模型提供依据。

4 依从性的行为预测指标

“预测未来行为的最佳指标是过去行为。”

- WalterMischel - B.F. Skinner - Mark Twain

正如删除线所示，这条格言在人类行为研究中流传已久。尽管经久不衰，它仍面临所有格言都同样面临的质疑：它真的具有普遍性吗？答案似乎是肯定的，但有一些前提：时间跨度较短、决策情境相似，以及决策者保持稳定 [26]。我们认为，过去行为格言适用于人机交互；判断未来是否会依从人工智能伙伴的建议，最佳方式是观察一次（或少数几次）交互并据此推断。

过去行为不仅仅是习惯的简单衡量指标。它反映了一系列能够稳定影响决策的因素。这会促使决策者保持自我一致性，至少部分解释了过去行为为何具有预测能力 [1,29]。这也表明，如果过去行为能够解释理论构念留下的残差，那么该构念就没有涵盖某些重要的预测因素。简而言之，该构念并不充分。仅凭这一点并不意味着该构念缺乏解释价值。例如，信任可能只是依从性的一小部分，而过去行为之所以能够预测未来服从，是因为它捕捉到了一些迄今尚未识别的因素。

我们认为，许多人机交互实验中用于预测未来服从的信任测量并非如此。这是因为这些信任测量（如 DTI）无法解释回归模型中相当大一部分方差。换句话说，它们不具备预测能力。此外，我们提出的替代预测指标——先前行为——则具备这种能力。还可以检验 DTI 或过去行为能否解释对方未能解释的数据方差。检验过去行为能否解释数据集中 DTI 遗留的残差很简单：只需在包含 DTI 的统计模型中加入过去行为这一控制变量，并观察残差是否显著减少。此外，还可以运用相同的基本思路，通过比较两个模型来判断 DTI 是否能够解释行为数据中有意义的方差：一个模型将 DTI 作为回归变量，另一个则不包含 DTI（也就是说，只包含实验处理控制变量）。

5 实证策略

我们分析了附录B中描述的三个数据集，以检验DTI和过去行为能否预测人机交互中的依从性，即未来行为（FB）。我们介绍了四种不同的过去行为测量指标。最简单的过去行为指标是参与者在实验环节中首次选择服从（FC）或忽略机器人建议的选择。研究1和研究2分别包含多个任务（三个和八个）。对于这两个数据集，我们将参与者在任务1中的依从性（M1C）作为过去行为的扩展测量指标。第三项研究仅包含一个较长的任务；不过，该研究中的每位参与者都沿着相同的轨迹行动，机器人也犯下了相同的一系列错误。因此，我们使用参与者在机器人犯下首次错误后的选择（AFM）中的依从性，以及截至AFM的平均依从性（AC-AFM）

来衡量过去行为。未来行为是各个模型的因变量，指参与者在剩余交互中服从机器人建议的次数百分比。因此，以FC为自变量的模型及用于比较这些模型的模型，使用的因变量（建筑物2的平均依从性n）不同于以M1C（任务1后的平均依从性）、AFM和AC-AFM（错误后建筑物的平均依从性错-误$+2$:n）为自变量的模型。请注意，其中两个数据集包含多个任务，这意味着各组建筑物之间存在间隔。我们的分析忽略了这一点，并将所有建筑物一视同仁。[3] 我们在附录B中详细介绍了数据和模型。

6 结果

如前所述，我们使用回归分析对数据进行建模。针对每个用例，我们报告的每个回归模型都以表格中的一列呈现，表格见附录。在报告回归模型的表格中，第一列始终是参照（空）模型。这些模型不包含 DTI 或先前行为测量，仅包含实验处理控制变量。处理条件的数值不在此报告，因为原始论文中已报告。对于连续变量，例如参与者的平均 DTI 响应，其系数应解释为：在其他条件不变的情况下，若该变量增加一个单位，因变量预期会发生的变化。对于指示变量，例如参与者在与机器人首次交互时是否依从，其系数值表示：在其他条件不变的情况下，自变量从 0 变为 1 时，因变量的平均变化。

6.1 研究1

信任倾向测量未能解释研究1数据中有意义的方差。在预测建筑物 2:24 的服从百分比时，由于模型中增加了一个回归变量而受到惩罚，调整后的 R^2 并不优于参照模型（见表 1 的第（1）列和第（2）列；所有表格均见附录 C）。不出所料，比较这两个模型的F检验拒绝了备择假设（$F = 0.832,\ \ p = 0.363$）。然而，加入 FC 测量（见表 1 第（3）列）后，模型解释的方差确实比参照模型更多（$F = 11.032,\ \ p = 0.001$）。同时包含两种测量的完整模型优于 DTI 模型（$F = 10.575,\ \ p = 0.001$），但不如 FC 模型（$F = 0.449,\ \ p = 0.504$）。

从使用一次交互来预测下一次，转变为使用任务一中所有的依从性决策来预测接下来两个任务中的决策

[3] 控制任务后，结果的解读并未发生实质性变化。

（16次交互）总体上带来了更好的模型拟合。实际上，这是一项合理性检查，因为我们使用参与者更多的过去行为来预测其较少的未来行为。结果模型见表3，并在表4中进行了比较，反映了功能连接测量指标的情况。M1C测量指标解释了有意义的方差，但DTI没有。总的来说，这些结果支持了以下假设：在研究1中，先前行为比DTI更能预测未来服从。

6.2 研究2

研究2的FC 测量结果与研究1的结果一致，如表5所示；不过需要注意的是，模型拟合总体较差（见表5中的 R^2 值）。DTI模型(2)并不比模型(1)更好，而模型(3)则成功解释了参数中显著更多的方差。再次将DTI测量加入模型(3)，并未改善模型的绩效。这意味着我们可以从研究2得出与研究1相同的结论：FC能够解释DTI无法解释的依从行为方差。M1C模型的拟合总体也较差（见表7中的 R^2 值）。此外，这些结果并未复现研究1的M1C模型的发现。目前尚不清楚这是否是研究设计差异或较低 n所导致的结果，还是确实反映了该测量的质量或可推广性。虽然我们无法断定先前行为在研究2中能够预测未来行为，但数据确实表明，先前行为比DTI更能预测未来行为。

6.3 研究3

研究3的结果比前两个实验更加耐人寻味。DTI 确实能够预测未来行为，但二者呈负相关。换言之，在 DTI 测量中得分较高的参与者，在整个实验过程中较少依从机器人的建议（见表9、11和13中的第（2）列）。模型拟合表明，DTI 测量得分每提高1个单位，预测依从性就会降低约5%的 ，这大致相当于比普通参与者多忽略机器人针对两栋建筑物提出的建议。这种效应也体现在F统计量 检验中：模型（2）解释的方差显著多于模型（1），因此，在预测对建筑物2:45的依从性时（$F = 8.885,\ \ p = 0.003$），以及在较短的时间范围9:45内预测依从性时（$F = 9.732,\ \ p = 0.002$），我们拒绝零假设。

这些模型还表明，FC、AFM 和 AC-AFM 指标能够有效预测未来的依从行为。在第一次交互中依从的参与者，在剩余任务中的依从程度高出近 12%（表 9 第 (3) 列）。使用机器人第一次犯错后作出的首次遵从决策时，也观察到了类似的效应（AFM；在这种情况下，依从的参与者在剩余任务中的依从程度高出近 19%；表 11 第 (3) 列）；使用机器人第一次犯错后首次遵从决策之前的平均依从程度时，也观察到了类似效应（AC-AFM；在这种情况下，

模型表明，完全不依从与完全依从的参与者之间，未来依从性相差约 31%；表 13 第 (3) 列）。

对于同时纳入 DTI 和既往行为测量的模型，在研究3中，我们实际上拒绝了“纳入 DTI 不会对仅包含行为测量的模型产生贡献”这一零假设（AF：$F = 8.985, p =$0.003，AFM：$F = 8.267, p = 0.005$，AC-AFM：$F = 9.046, p = 0.003$）。重要的是，DTI 的预测效应仍然与依从性呈负相关，其影响程度大致相同。总而言之，我们在这三项研究中提出的模型带来了两点启示：第一，过去的依从行为，即使形式最为简单，也是未来依从性更可靠、更一致的预测因素；第二，DTI 的可靠性较低，而且可能呈现出与文献假设相反的负相关。

6.4 另一种解释

McKnight 等人 [23] 明确指出，DTI 应与早期服从呈正相关，而不一定与总体服从相关。DTI 可能确实能够预测参与者的首次服从选择（FC）。这意味着，与其比较 FC 和 DTI，不如使用 DTI 来预测 FC。我们针对每项研究使用多元线性回归检验了这一假设，并在表 15 中报告了结果。我们还在每个模型中纳入了处理条件控制变量；与其他表格一样，其存在通过勾选标记表示，相关的系数值和截距则予以删略以作简化。总而言之，在任何一项研究中，DTI 都未能显著预测参与者的首次服从选择。我们将此视为有力证据，支持“DTI 与早期服从选择不相关”这一零假设。我们还扩大了早期服从的定义，将前五次服从选择纳入其中；然而，结果仍然相同，因此我们没有将结果作为附加表格列出。

7 讨论

我们对以往采用信任倾向量表的研究进行了综述和实证评估，由此得出结论：DTI 并不能准确预测人机交互中的依从性。此前三项实验的数据支持这一结论。此外，我们采用的基本方法可作为评估其他测量工具的范例。

在我们获取数据的各项研究中，机器人预测何时需要防护装备的表现显著优于随机猜测（例如，即使能力较低的机器人，预测准确率也达到 80%）。这意味着依从性与任务成功高度相关。正如早期依从行为能够预测后期服从一样，数据也表明，早期行为可以预测任务成功。当然，据此得出结论并不妥当，因为这只是相关性造成的假象。我们认为，有必要更深入地研究早期行为及其与

后期服从之间的关系。深入理解这种关系或许有助于设计依从性干预，即使它并不适用于所有情境（显然，一次性的交互无法从对早期行为的洞见中获益）。

如果 DTI 在模型中能够解释更多显著方差，我们就可以检验它是否捕捉到了未来行为测量所解释的方差，反之亦然。遗憾的是，我们现有的数据集无法支持此类分析。更大规模的数据集，或许来自 DTI 预测价值更高的不同情境，可能有助于研究人员区分各项测量所捕捉到、影响依从性的因素。

值得指出的是，信任与依从性并非完全一致的概念。在人机交互和 HRI 领域，人们通常会引用 Lee 和 See 对信任的定义：在具有不确定性和脆弱性的情境中，相信智能体会帮助实现个人目标的态度（第 51 [21] 页）。这个简单的定义掩盖了人类认知中一种极其复杂的现象：人们对对象及其属性所形成的态度。另一方面，依从性相对简单：一个人是否（恰当地）遵从了自动化系统给出的建议？关于信任与依从性的文献中，一个常见假设是，人类会对自动化系统及其各种属性持有一种特定且可识别的态度（信任）；这种态度适用范围广泛，包括预测依从性。此外，该假设认为，这种态度可以通过简短的问题量表（例如 DTI）轻松测量。一般而言，人类持有的态度一向难以研究；我们认为，态度的这一特征使此类方法容易失效，尤其是在所关注的行为与特定测量指标没有直接联系时。

在我们提取数据的实验中，机器人是默认的主体，而其预测危险的可靠性则是隐含的关注属性。若要支持标准假设，至少需要满足以下几点：

– 信任倾向假设认为，信任在某种可测量的程度上，是认知中类似特质的一面。– 信任倾向可通过 DTI 等量表准确测量。– 信任倾向普遍适用于主体及其属性，即不存在例外情况，也没有会中介或调节这种关系的变量。– 依从性与信任倾向呈正相关。

如果其中任何一点不成立，我们就会预期 DTI 对依从性的预测表现不佳，甚至完全失效；而这恰恰是我们观察到的结果。遗憾的是，现有数据不足以让我们进一步探究我们究竟在何种意义上否定了标准假设。

Lee 和 See 指出，信任是一种多维构念，并且甚至表示：“整体信任与特定信任之间似乎没有可靠的关系……要使对自动化的信任具有高度的功能特异性，可能需要针对

自动化的每个细节层级提供具体信息。”（第 58 [21] 页）。这一见解本身就足以动摇上述第一、第二和第三项，而它也得到了认知科学大量研究的支持。承认人们对不同事物所持的想法和态度长期以来一直是科学探究的主题，且众所周知难以达成共识，更进一步削弱了这三项。历史上不乏试图驾驭这些问题的著名哲学努力。从亚里士多德在《解释篇》 [39] 中的探索，到丹尼尔·丹尼特的《意向立场》[12] 和露丝·米利肯的生物语义学 [25] ，其中不乏值得称道的尝试。然而，关于对某个对象持有一种态度究竟意味着什么，仍有许多未解之问。现代心理学同样深受困扰：尽管布伦塔诺曾试图通过对意向性的研究来厘清这一问题 [9], ，并对从西格蒙德·弗洛伊德到伯特兰·罗素的德国经验心理学传统产生了深远影响，但这仍是一个争论激烈的话题 [10]。

最后一点，即依从性与信任呈正相关，同样存在诸多问题。近期疫情提供了一个恰当的案例研究：研究发现，在疫情初期，信任倾向与欧洲公民对政府政策的依从性显著正相关 [7], 然而，在新加坡，尽管民众普遍信任政府，即使政府努力改变公民的风险认知，依从性依然不佳 [47]。综合来看，这些发现令人质疑依从性与信任之间相关性的可靠性。总而言之，结合我们的分析，我们认为标准假设在多个方面都站不住脚：它依赖类似信任的特质，因而同时过于宽泛；又由于忽视了重要的混杂因素，而过于具体。

尽管如此，DTI 等测量指标仍可能具有价值，尤其是在它们被证明具有预测价值时（无论相关性的方向如何）。如果某项测量指标能够可靠地解释人类行为的方差，那么它在开发自动化系统时就有价值，因为解释方差是预测选择的核心。据我们所知，我们的分析并未驳斥任何人机交互论文中的发现。我们的核心结论是，我们不仅应通过全面的文献综述审查测量指标，还应针对人类与人工智能交互领域所需的特定建模，对这些指标进行验证。

我们对社区的总体建议如下：

– **针对特定模型和结果测量，验证从其他学科借鉴的心理测量指标**。此外，正如 [36], 所指出的，从李克特量表中只采用单个条目（或部分条目）会削弱量表原有的信度和效度。– **考虑采用可能已存在于数据中的其他简易测量方法**。这些测量方法不一定要取代其他方法，也可以作为有价值的补充，从而得出更稳健的推断。– **评估新测量方法相较于其他方法的效用**。通过统计分析了解不同测量方法解释数据方差的方式，不仅能更严谨地考察这些替代性原因，还能带来新的见解并推动新的建模方案。

如果我们在数据中观察到的效应仅存在于 DTI、信任和依从性中，那将令人难以置信。人类对自动化系统抱有的其他心理状态（如愤怒、信念、恐惧、希望等）及其相关结果，很可能也存在类似效应。例如，大量研究表明，对机器人的负面态度能够高度预测交互及其结果 [28,27]。我们推测，正如早期服从比 DTI 更能预测后期服从一样，也存在一些行为测量，其表现优于对机器人的负面态度测量工具，和/或侵扰性更低。我们认为，识别、检验并验证此类测量方法，将是改善人类与各类自动化系统（从推荐系统到机器人）之间关系的重要一步。

8 结论

用于测量不同心理学构念的量表在为人机交互模型提供信息方面十分常见，但也经常被误用 [36]。在本文中，我们评估了其中一种量表的使用情况。此前的研究使用该量表，希望它能够预测重要的人类行为（如依从性）以及总体结果（如任务表现）。该量表在两项实验中未能发挥预测作用，并且在第三项实验中呈现出与量表编制者所预测的相反相关性。我们发现，一项简单的行为测量指标能够更好地预测未来服从，因此也将是依从性推理模型中更有信息价值的输入。本文用于评估这两种测量指标的方法十分直接，可作为其他研究人员评估替代测量指标的原型，而这些指标可以作为人机交互模型的输入。

参考文献

1. Ajzen, I.：《计划行为理论》。《组织行为与人类决策过程》**50**(2)，179–211（1991） 2. Aliasghari, P.、Ghafurian, M.、Nehaniv, C.L.、Dautenhahn, K.：家用训练机器人犯错对人类教师信任的影响。载于：IEEE机器人与人类交互通信国际会议（RO-MAN）论文集。第81–88页。IEEE（2021） 3. Aliasghari, P.、Ghafurian, M.、Nehaniv, C.L.、Dautenhahn, K.：犯错的学生机器人采用不同口头表达方式时，会如何影响人类教师使用该机器人的意图？载于：第9届人类—智能体交互国际会议论文集。第21–30页（2021） 4. Amershi, S.、Weld, D.、Vorvoreanu, M.、Fourney, A.、Nushi, B.、Collisson, P.、Suh, J.、Iqbal, S.、Bennett, P.N.、Inkpen, K.等：人机交互指南。载于：2019年CHI计算机系统中的人为因素会议论文集。第1–13页（2019） 5. Aroyo, A.M.、Rea, F.、Sandini, G.、Sciutti, A.：人机交互中的信任与社会工程：机器人会让你披露敏感信息、遵从其建议或进行赌博吗？《IEEE机器人与自动化快报》**3**(4)，3701–3708（2018）

6. Ashleigh, M.J.、Higgs, M.、Dulewicz, V.：新的信任倾向量表及其与个体福祉的关系：对人力资源管理政策与实践的启示。《人力资源管理期刊》**22**(4)，360–376（2012年） 7. Bargain, O.、Aminjonov, U.：新冠肺炎时期对公共卫生政策的信任与依从性。《公共经济学期刊》**192**，104316（2020） 8. Barnes, M.J.、Wang, N.、Pynadath, D.V.、Chen, J.Y.：人机双向透明度。载于：《人机交互中的信任》，第209–232页。Elsevier（2021） 9. Brentano, F.：《经验立场上的心理学》。Routledge（2012年） 10. Chater, N.、Zeitoun, H.、Melkonyan, T.：社会交互的悖论：共享意向性、我们推理与虚拟谈判。《心理学评论》**129**(3)，415（2022年） 11. Chi, O.H.、Jia, S.、Li, Y.、Gursoy, D.：开发用于测量消费者对服务交付中与人工智能（AI）社交机器人交互之信任的形成性量表。《人类行为中的计算机》**118**，106700（2021） 12. Dennett, D.C.：《意向立场》。麻省理工学院出版社（1987年） 13. Elliot, J.：面向成功团队的人工社交智能（ASIST）（2021），https://www.darpa.mil/program/artificial-social-intelligence-for-successful-teams 14. Gurney, N.、Pynadath, D.、Wang, N.：我的行动比你的言语更响亮：用户行为何时能够预测其对智能体属性的信念。arXiv预印本 arXiv:2301.09011（2023） 15. Gurney, N.、Pynadath, D.V.、Wang, N.：测量和预测人类对人工智能队友建议的信任。载于：人机交互国际会议。第22–34页。Springer（2022） 16. Hancock, P.A.、Billings, D.R.、Schaefer, K.E.、Chen, J.Y.、De Visser, E.J.、Parasuraman, R.：影响人机交互中信任的因素：一项荟萃分析。《人为因素》**53**(5)，517–527（2011） 17. Hoff, K.A.、Bashir, M.：对自动化的信任：整合影响信任的因素的经验证据。《人为因素》**57**(3)，407–434（2015年） 18. Jessup, S.A.、Schneider, T.R.、Alarcon, G.M.、Ryan, T.J.、Capiola, A.：自动化信任倾向的测量。载于：人机交互国际会议。第476–489页。Springer（2019） 19. Kaelbling, L.P.、Littman, M.L.、Cassandra, A.R.：在部分可观测随机领域中进行规划与行动。《人工智能》**101**(1)，99–134（1998） 20. Kaelbling, L.P.、Littman, M.L.、Moore, A.W.：强化学习：一项调查。《人工智能研究期刊》**4**，237–285（1996） 21. Lee, J.D.、See, K.A.：对自动化的信任：为适当依赖而设计。《人为因素》**46**(1)，50–80（2004） 22. Lutz, C.、Tam´o-Larrieux, A.：机器人隐私悖论：理解隐私担忧如何塑造使用社交机器人的意图。《人机传播》**1**，87–111（2020） 23. McKnight, D.H.、Choudhury, V.、Kacmar, C.：开发并验证用于电子商务的信任测量：一种整合型分类体系。《信息系统研究》**13**(3)，334–359（2002） 24. Merritt, S.M.、Huber, K.、LaChapell-Unnerstall, J.、Lee, D.：自动化系统信任的持续校准。技术报告，密苏里大学圣路易斯分校（2014） 25. Millikan, R.G.：生物语义学。《哲学杂志》**86**(6)，281–297（1989） 26. Mischel, W.：人格与评估。Psychology Press（2013） 27. Nomura, T.、Kanda, T.、Suzuki, T.：对机器人持负面态度如何影响人机交互的实验研究。《人工智能与社会》**20**(2)，138–150（2006）

28. Nomura, T., Suzuki, T., Kanda, T., Kato, K.：负面机器人态度的测量。《交互研究》**7**(3), 437–454 (2006) 29. Ouellette, J.A., Wood, W.：日常生活中的习惯与意图：过去行为预测未来行为的多重过程。《心理学通报》**124**(1), 54 (1998) 30. Parasuraman, R., Riley, V.：人类与自动化：使用、误用、弃用与滥用。《人为因素》**39**(2), 230–253 (1997) 31. Pynadath, D.V., Gurney, N., Wang, N.：人机团队中的可解释强化学习：决策树解释对透明度的影响。载于：2022年第31届IEEE机器人与人类交互通信国际会议（RO-MAN）。第749–756页。IEEE (2022) 32. Pynadath, D.V., Wang, N., Kamireddy, S.：一种预测人类—智能体交互中信任行为的马尔可夫方法。载于：第七届人类—智能体交互国际会议论文集。第171–178页 (2019) 33. Rossi, A., Dautenhahn, K., Koay, K.L., Walters, M.L.：个人倾向与人格特质对人们在紧急情境中信任机器人的影响。《Paladyn：行为机器人学杂志》**9**(1), 137–154 (2018) 34. Rossi, A., Dautenhahn, K., Koay, K.L., Walters, M.L., Holthaus, P.：在重复交互研究中评估人们对机器人信任的感知。载于：国际社会机器人会议。第453–465页。Springer (2020) 35. Schaefer, K.：人机信任的感知与测量 (2013), https://stars.library.ucf.edu/etd/2688 36. Schrum, M.L., Johnson, M., Ghuy, M., Gombolay, M.C.：回顾四年：人机交互研究中李克特量表的统计实践。载于：2020年ACM/IEEE人机交互国际会议伴会。第43–52页 (2020) 37. Seeber, I., Bittner, E., Briggs, R.O., De Vreede, T., De Vreede, G.J., Elkins, A., Maier, R., Merz, A.B., Oeste-Reiß, S., Randrup, N., 等：作为队友的机器：团队协作中人工智能的研究议程。《信息与管理》**57**(2), 103174 (2020) 38. Shneiderman, B.：以人为中心的人工智能：可靠、安全且值得信赖。《国际人机交互杂志》**36**(6), 495–504 (2020) 39. Stevenson, D.C.：互联网经典文库：亚里士多德《解释篇》 (2009), http://classics.mit.edu/Aristotle/interpretation.html 40. Sutton, R.S., Barto, A.G.：《强化学习：导论》。麻省理工学院出版社 (2018) 41. Tauchert, C., Mesbah, N., 等：跟随机器人？调查用户对机器人顾问建议的采纳情况。载于：信息系统国际会议论文集 (2019) 42. Textor, C., Pak, R.：关注信任：探究注意力控制与对自动化的信任之间的关系。载于：人因与工效学学会年会论文集。第65卷第1期，第817–821页。SAGE出版社，加利福尼亚州洛杉矶 (2021) 43. Venkatesh, V.：感知易用性的决定因素：将控制、内在动机和情绪融入技术接受模型。《信息系统研究》**11**(4), 342–365 (2000) 44. Wang, N., Pynadath, D.V., Hill, S.G.：基于POMDP生成的解释对人机团队信任与绩效的影响。载于：2016年自主智能体与多智能体系统国际会议论文集。第997–1005页 (2016)

45. Wang, N., Pynadath, D.V., Hill, S.G.：人机团队中的信任校准：比较自动生成的解释。载于：2016年第11届ACM/IEEE人机交互国际会议（HRI）。第109–116页。IEEE (2016) 46. Wang, N., Pynadath, D.V., Rovira, E., Barnes, M.J., Hill, S.G.：是我的外表，还是我说的话？解释、具身化与预期对人机团队信任与绩效的影响。载于：说服技术国际会议。第56–69页。Springer (2018) 47. Wong, C.M.L., Jensen, O.：信任悖论：新冠肺炎疫情期间新加坡的感知风险与公众遵从。《风险研究杂志》**23**(7-8), 1021–1030 (2020) 48. Xu, A., Dudek, G.：Optimo：用于非对称人机协作的在线概率信任推断模型。载于：2015年第10届ACM/IEEE人机交互国际会议（HRI）。第221–228页。IEEE (2015) 49. Yagoda, R.E., Gillan, D.J.：你希望我信任机器人吗？人机交互信任量表的开发。《国际社会机器人学杂志》**4**(3), 235–248 (2012)

附录 A 图

Fig。**1.** 迪sposition to Trust Inventory Items

仁慈。

1 一般来说，人们确实关心他人的福祉。2. 一般人真心关心他人的问题。3. 大多数时候，人们足够关心他人，愿意尽力提供帮助，而不是只顾自己。

诚信。

1 总的来说，大多数人都会信守承诺。2. 我认为，人们通常会努力做到言行一致。3. 大多数人在与他人交往时都很诚实。

能力。

1 我相信，大多数专业人士都能出色地完成工作。2. 大多数专业人士在自己选择的领域都非常博学。3. 绝大多数专业人士在各自的专业领域都具备胜任能力。

信任倾向。

1 我通常会信任别人，除非他们给我不信任他们的理由。2. 初次见面时，我通常会先对别人抱持善意。3. 我一贯的做法是信任新认识的人，直到他们证明我不该信任他们。

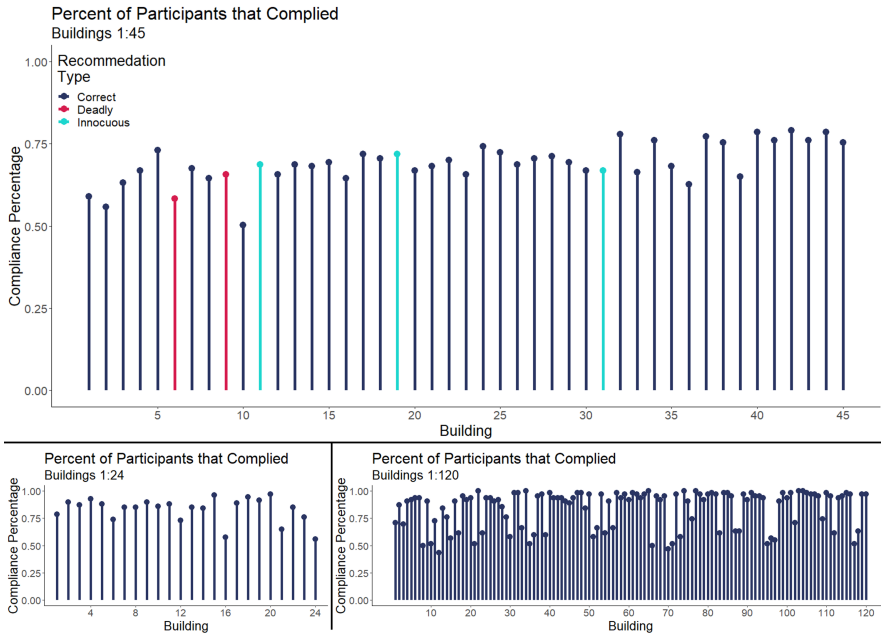


图2. 机器人只在研究3中改变了沟通方式——它始终以相同的顺序犯相同的错误。这些特点使数据适合以简洁、直观的方式进行可视化呈现，我们在上方面板中展示了这种呈现方式。请注意，该面板中的颜色表示机器人的建议类型。具体而言，机器人犯了两个致命错误和三个无害错误，这些错误已在上方突出显示。总体而言，依从性呈上升趋势。研究1和研究2的图分别位于左下和右下方面板，仅显示服从百分比，而未显示机器人所提出建议的类型。

附录B 数据与模型

B.1 模型

我们使用线性回归对各种行为测量进行建模，并检验其预测价值。基本方法是拟合一个参照模型（**模型1**），仅根据处理条件预测结果测量——未来行为，例如：

$$Y_{iFB} = \beta_0 + \beta_{\text{Treat}} X_{\text{iTreat}} + \epsilon_i \tag{1}$$

我们在各项实验中都采用这种一般形式的参照（零）模型。请注意，FB 测量有多种；在一组特定模型中使用哪一种，取决于所讨论的自变量。不过，给定的参照模型和检验模型始终使用相同的因变量。

模型2：第一类检验模型将参与者的DTI得分纳入自变量：

$$Y_{iFB} = \beta_0 + \beta_{\text{Treat}} X_{\text{iTreat}} + \beta_{\text{DTI}} X_{\text{iDTI}} + \epsilon_i \tag{2}$$

请注意，参照模型嵌套在此模型之中。换言之，方程（2）代表备择假设：将预测变量添加到模型中将能够解释更多方差，或显著提升模型绩效。比较这两个模型很简单，只需进行F检验；在此情形下，该检验可以判断更复杂的模型是否会产生统计上不同的残差平方和值。如果是，我们便可以拒绝零假设（即参照模型已足够），转而支持备择假设，也就是说，添加额外的预测变量是合理的。在此例中，F检验可用于考察在模型中添加 DTI 或其他测量指标是否有用。[4] 当检验结果的 p值小于 0.05 时，我们可以得出结论：两个模型的残差平方和值在 $\alpha = 0.05$水平上存在显著差异。换言之，新的解释变量是合理的，因为它显著增加了模型所解释的方差量（即 RSS 更低）。

模型3：下一类检验模型将参与者的过去行为作为自变量纳入，例如以下模型：

$$Y_{iFB} = \beta_0 + \beta_{\text{Treat}} X_{\text{iTreat}} + \beta_{\text{FC}} X_{\text{iFC}} + \epsilon_i \tag{3}$$

本模型使用参与者的首次选择，来预测他们在其余所有选择中的依从性。同样，该模型包含参照模型，简单的F检验即可揭示FC预测变量的效用。我们为M1C、AFM和AC-AFM构建了类似的模型（但使用相应的FB测量）。

[4] 似然比检验是比较此类模型的另一种方法，也能得出类似的见解。

模型4：由于DTI和既往行为测量很可能在模型中解释不同的方差，因此直接通过 R^2 比较方程（2）和（3）并不能充分说明问题。因此，我们引入第三类测试模型，其中同时纳入DTI和一项既往行为测量。对于FC，模型如下：

$$Y_{iFB} = \beta_0 + \beta_{\text{Treat}} X_{\text{iTreat}} + \beta_{\text{DTI}} X_{\text{iDTI}} + \beta_{\text{FC}} X_{\text{iFC}} + \epsilon_i \tag{4}$$

这些模型有助于评估同时纳入 DTI 和过去行为测量是否值得增加模型复杂度；同样，我们使用*F*统计量检验。最后，为便于阅读，我们在论文正文中提及相关统计量，但将每组模型及其相关检验的回归表和其他表格置于附录中。这些表格按上述相同顺序列出，有助于比较 DTI 与行为测量的模型。每张回归表后均附有一张表，呈现相关*F*统计量检验的结果。

B.2 数据

我们使用了一个长期研究项目中开展的实验数据，该项目研究可解释性与人工智能。参与这些实验的参与者在侦察任务中与模拟机器人组队。任务要求进入建筑物，以确定是否存在威胁。机器人先行，并配备了摄像机、麦克风以及用于探测核、生物和化学威胁的传感器。这些传感器并非完全可靠。机器人根据使用传感器收集的数据，向参与者提出是否穿戴防护装备的建议。随后，参与者决定是否穿戴装备，即是否遵从机器人的建议。参与者穿戴装备时，装备总能消除任何威胁。如果他们不穿戴装备却遭遇威胁，就会在虚拟世界中死亡；但在现实中，他们会受到严厉的时间惩罚。最后，参与者穿戴装备时会遭遇轻微的时间延迟（远短于在虚拟世界中死亡所造成的延迟）。

在这三项研究中，机器人都将带有噪声的传感器读数作为输入，并依据通过部分可观测马尔可夫决策过程（POMDP） [19] 或无模型强化学习（RL） [20,40] 计算出的策略，给出建议；奖励信号则由时间成本和所造成的死亡人数构成。在所有研究中，机器人的表现都显著优于随机猜测。这意味着，参与者的依从性与其是否做出规范性选择高度相关，也就是说，是否在正确的时间穿戴防护装备。

在开始各自的任务之前，三项研究的参与者都完成了包含 12 个条目的 DTI。有关这些实验条件的总体详情和结果，已在原始论文中报告。为求简洁，我们在此不再重复这些发现。不过请注意，由于观测值不完整（部分参与者选择不完成 DTI），我们报告的 n可能与原始论文中的数值有所不同。

研究1的参与者（$n = 198$，亚马逊土耳其机器人）完成了三项任务，每项任务涉及八栋建筑物 [44]。他们被随机分配，与以下其中一种

基于部分可观测马尔可夫决策过程的机器人类型配对：一种能力较强、从不出错的机器人，或一种能力较弱、20% 的情况下会出错（即可靠性为 80%）的机器人。两种机器人类型分别与四种不同的建议解释条件交叉组合：无解释、置信度、传感器读数版本1和传感器读数版本2。这项实验完全采用被试间设计，也就是说，每位参与者在整个任务过程中只与一种机器人类型互动，并且只接收一种类型的解释。研究1模型中的系数 β_{Treat} 表示参与者所经历的信息条件。首次服从选择（FC）在参与者采纳机器人对第一栋建筑的建议时取值为1，否则取值为0。另一方面，任务1服从率（M1C）是参与者在第一项任务中服从机器人建议的次数比例。因而，研究1中与FC相关的后续服从行为（FB）指标，是参与者在其余23栋建筑上服从建议的次数比例。同样，对于M1C，该指标是参与者在第二和第三项任务中听从机器人建议的次数比例。需要注意的是，参与者并未被告知自己在各项任务中是否与同一机器人互动；相反，机器人每次开始一项任务时，都会表现得仿佛此前从未与该参与者互动过。

研究2的参与者（$n = 53$，西点军校的学员）完成了八项任务，每项任务都配有一台不同的基于部分可观测马尔可夫决策过程的机器人 [46]。在每项任务中，人机团队对15栋建筑物执行侦察任务。任务顺序固定（即始终以相同顺序搜索建筑物，且各项任务之间的顺序相同），但机器人的顺序随机安排。$2\times2\times2$ 设计交叉考察了机器人是否承认错误（不承认/承认）、建议的解释方式（无/置信度）和具身化（类机器人/类狗）。与研究1不同，参与者在每项任务中都会与不同的机器人互动。尽管如此，为了证明简单行为测量的稳健性，我们仍采用与研究1相同的首次服从选择（FC）以及类似的任务1服从率（M1C）。显然，这些依从性测量涵盖了更长的时间范围：分别涉及119项任务和105项任务。研究2数据的模型 β_{Treat} 纳入了首项任务中机器人的类型。机器人顾问的顺序可能会产生影响；然而，数据不足以构建能够揭示这一特征的层级模型。

研究3的参与者（$n = 148$，亚马逊土耳其机器人）与一个基于强化学习（RL：强化学习）的机器人完成了一项涵盖45栋建筑物的任务，采用完全组间设计 [31,15,14]。各处理条件下机器人的能力保持不变，但对其建议的解释方式有所不同：不提供解释、解释决策、解释决策与学习。同前两项研究一样，首次服从选择（FC）的定义相同，而FC的结果测量是参与者在剩余44栋建筑物中的服从比例。由于整个实验只包含一项任务，因此任务1服从率（M1C）不适用。不过，由于各处理条件下建筑物的顺序和机器人的绩效均固定，另外两项服从率测量——首次错误后的选择（AFM）和首次错误前的平均服从率（AC-AFM）——便具有意义。首次错误发生在

第六栋建筑物，因此，参与者对第七栋建筑物的决定即为AFM测量值，而他们在前七栋建筑物中服从建议的比例即为AC-AFM。相应地，因变量是某位参与者在剩余38栋建筑物中服从建议的比例。

附录 C 表格

表 1. 研究1 功能连接模型

	Dependent variable: Compliance Percentage Buildings 2:24			
	回归模型			
	(1)	(2)	(3)	(4)
DTI		0.007		0.005
		(0.008)		(0.008)
FC			0.086**	0.084**
			(0.026)	(0.026)
处理控制变量	✓	✓	✓	✓
常数项	0.762***	0.724***	0.687***	0.661***
	(0.032)	(0.052)	(0.038)	(0.054)
观测值	198	198	198	198
R^2	0.420	0.422	0.452	0.453
调整后 R^2	0.398	0.398	0.429	0.427
残差标准误差	0.152（自由度 = 190）	0.152（自由度 = 189）	0.148（自由度 = 189）	0.148（自由度 = 188）
F统计量	19.639***（自由度 = 7; 190）	17.273***（自由度 = 8; 189）	19.471***（自由度 = 8; 189）	17.307***（自由度 = 9; 188）
Note:				*p<0.05; **p<0.01; ***p<0.001

表2. 研究1 功能连接模型比较

统计量	(1) 与 (2)	(1) 与 (3)	(2) 与 (4) 对比	(3) 与 (4) 对比
平方和	0.019	0.242	0.232	0.010
F	0.832	11.032	10.575	0.449
Pr(>F统计量)	0.363	0.001	0.001	0.504

表3。 研究1 M1C模型

	Dependent variable: Compliance Percentage Buildings 9:24			
	回归模型			
	(1)	(2)	(3)	(4)
DTI		0.003		−0.0004
		(0.009)		(0.008)
M1C			0.242***	0.243***
			(0.052)	(0.052)
处理控制变量	✓	✓	✓	✓
常数项	0.770***	0.754***	0.587***	0.589***
	(0.035)	(0.058)	(0.052)	(0.066)
观测值	198	198	198	198
R^2	0.451	0.451	0.508	0.508
调整后的 R^2	0.431	0.428	0.487	0.484
残差标准误	0.170	0.170	0.161	0.161
	（自由度 = 190）	（自由度 = 189）	（自由度 = 189）	（自由度 = 188）
F统计量	22.282***	19.421***	24.384***	21.561***
	（df = 7；190）	（df = 8；189）	（df = 8；189）	（自由度 = 9；188）
Note:				*p<0.05；**p<0.01；***p<0.001

表4. 研究1 M1C模型比较

统计量	(1) 与 (2)	(1) 与 (3)	(2) 与 (4)	(3) 与 (4)
平方和	0.003	0.568	0.565	<0.001
F	0.119	21.923	21.677	0.002
Pr(>F统计量)	0.731	<0.001	<0.001	0.965

表5。研究2功能连接模型

	Dependent variable: Compliance Percentage Buildings 2:120			
	回归模型			
	(1)	(2)	(3)	(4)
DTI		0.004		0.004
		(0.011)		(0.011)
FC			0.043*	0.043*
			(0.018)	(0.018)
处理控制变量	✓	✓	✓	✓
观测值	49	49	49	49
R^2	0.047	0.050	0.163	0.165
调整后的 R^2	−0.016	−0.037	0.087	0.068
残差标准误	0.053	0.053	0.050	0.051
	（自由度 = 45）	（自由度 = 44）	（自由度 = 44）	（自由度 = 43）
错误				
F统计量	0.742	0.574	2.137	1.704
	（自由度 = 3; 45）	（自由度 = 4; 44）	（自由度 = 4; 44）	（df = 5; 43）
Note:				*p<0.05; **p<0.01; ***p<0.001

表6. 研究2功能连接模型比较

统计量	(1) 与 (2) 比较	(1) 与 (3) 比较	(2) 与 (4) 比较	(3) 与 (4) 比较
平方和	<0.001	0.015	0.015	<0.001
F	0.112	6.071	5.964	0.137
Pr(>F统计量)	0.739	0.017	0.019	0.713

表 7. 研究2 M1C模型

	Dependent variable: Compliance Percentage Buildings 16:120			
	回归模型			
	(1)	(2)	(3)	(4)
DTI		0.003		0.004
		(0.012)		(0.013)
M1C			−0.015	−0.016
			(0.067)	(0.068)
处理控制变量	✓	✓	✓	✓
观测值	49	49	49	49
R^2	0.041	0.042	0.042	0.043
调整后 R^2	−0.023	−0.045	−0.045	−0.068
残差标准误	0.059	0.060	0.060	0.060
	（自由度 = 45）	（自由度 = 44）	（自由度 = 44）	（自由度 = 43）
F统计量	0.634	0.485	0.478	0.390
	（自由度 = 3; 45）	(df = 4; 44)	(df = 4; 44)	(df = 5; 43)
Note:				*p<0.05; **p<0.01; ***p<0.001

表8。 研究2 M1C模型比较

统计量	(1) 与 (2) 对比	(1) 与 (3) 对比	(2) 与 (4) 对比	(3) 与 (4) 对比
平方和	<0.001	<0.001	<0.001	<0.001
F	0.076	0.051	0.054	0.078
Pr(>F统计量)	0.785	0.822	0.818	0.782

表9. 研究3功能连接模型

	Dependent variable: Compliance Percentage Buildings 2:45			
	回归模型			
	(1)	(2)	(3)	(4)
DTI		−0.052**		−0.050**
		(0.017)		(0.017)
FC			0.115***	0.112***
			(0.031)	(0.030)
处理控制变量	✓	✓	✓	✓
观测值	148	148	148	148
R^2	0.040	0.096	0.125	0.177
调整后 R^2	0.027	0.077	0.107	0.154
残差标准误	0.193	0.188	0.185	0.180（自由度 = 143）
	（自由度 = 145）	（自由度 = 144）	（自由度 = 144）	
错误				
F统计量	3.011	5.078**	6.850***	7.669***
	（自由度 = 2; 145）	（自由度 = 3; 144）	（自由度 = 3; 144）	（自由度 = 4; 143）
Note:				*p<0.05; **p<0.01; ***p<0.001

表10. 研究3 功能连接模型比较

统计量	(1) 与 (2) 比较	(1) 与 (3) 比较	(2) 与 (4) 比较	(3) 与 (4) 比较
平方和	0.313	0.477	0.454	0.290
F	8.885	12.988	14.058	8.985
Pr(>F统计量)	0.003	<0.001	<0.001	0.003

表 11. 研究3 AFM模型

	Dependent variable: Compliance Percentage Buildings 8:45			
	回归模型			
	(1)	(2)	(3)	(4)
DTI		−0.056**		−0.047**
		(0.018)		(0.016)
AFM			0.187***	0.179***
			(0.032)	(0.032)
处理控制变量	✓	✓	✓	✓
观测值	148	148	148	148
R^2	0.037	0.098	0.220	0.263
调整后 R^2	0.024	0.080	0.204	0.242
残差标准误差	0.200（自由度 = 145）	0.194（自由度 = 144）	0.181（自由度 = 144）	0.176（自由度 = 143）
F统计量	2.822（自由度 = 2; 145）	5.238**（自由度 = 3; 144）	13.577***（自由度 = 3; 144）	12.763***（自由度 = 4; 143）
Note:				*p<0.05; **p<0.01; ***p<0.001

表12. 研究3 AFM模型比较

统计量	(1) 与 (2) 比较	(1) 与 (3) 比较	(2) 与 (4) 比较	(3) 与 (4) 比较
平方和	0.367	1.103	0.991	0.256
F	9.732	33.811	31.960	8.267
Pr(>F统计量)	0.002	<0.001	<0.001	0.005

表13. 研究3 AC-AFM模型

	Dependent variable: Compliance Percentage Buildings 8:45			
	回归模型			
	(1)	(2)	(3)	(4)
DTI		−0.056**		−0.049**
		(0.018)		(0.016)
AC-AFM			0.313***	0.301***
			(0.053)	(0.052)
处理控制变量	✓	✓	✓	✓
观测值	148	148	148	148
R^2	0.037	0.098	0.223	0.269
调整后的 R^2	0.024	0.080	0.207	0.249
残差标准误差	0.200 (df = 145)	0.194 (自由度 = 144)	0.180 (自由度 = 144)	0.175 (自由度 = 143)
F统计量	2.822 (自由度 = 2; 145)	5.238** (自由度 = 3; 144)	13.767*** (自由度 = 3; 144)	13.164*** (自由度 = 4; 143)
Note:				*p<0.05; **p<0.01; ***p<0.001

表14. 研究3 AC-AFM模型比较

统计量	(1) 与 (2) 比较	(1) 与 (3) 比较	(2) 与 (4) 比较	(3) 与 (4) 比较
平方和	0.367	1.116	1.028	0.278
F	9.732	34.359	33.403	9.046
Pr(>F统计量)	0.002	<0.001	<0.001	0.003

表 15. DTI 作为首次服从选择的预测因素

	Dependent variable: First Choice		
	回归模型		
	研究1	研究2	研究3
DTI	0.024	0.002	−0.017
	(0.021)	(0.092)	(0.047)
控制变量	✓	✓	✓
观测值	198	48	148
R^2	0.062	0.145	0.003
调整后的 R^2	0.022	0.065	−0.018
残差标准误	0.415	0.434	0.499
	(自由度 = 189)	(自由度 = 43)	(自由度 = 144)
F统计量	1.562	1.820	0.141
	（自由度 = 8; 189）	（自由度 = 4; 43）	（自由度 = 3; 144）
Note:			*p<0.1; **p<0.05; ***p<0.01