arXiv:2302.10088v1 [cs.MM] 20 Feb 2023

注册报告：基于他人个人价值观对其音乐偏好的感知

Sandy Manolios*、Catholijn M. Jonker、Cynthia C.S. Liem

荷兰代尔夫特理工大学智能系统系

* s.manolios@tudelft.nl

摘要

本研究是一个研究系列的一部分，该系列旨在揭示人们音乐偏好背后的奥秘，以便提供更好的音乐推荐。更具体地说，本研究从个人价值观的角度切入。个人价值观是人们所追求的目标，也是营销研究中广泛用于了解客户对特定产品类型偏好的工具。因此，探究个人价值观在音乐领域的适用性是有意义的。基于一项采用手段—目的理论的先前定性研究，我们设计了一份调查问卷，尝试以更定量的方式探讨个人价值观与音乐偏好之间的关系。我们还借助模拟研究支持这一研究方法，以改进实验程序和相关决策。

1 引言

人人都喜欢音乐，但世界上存在种类繁多、风格迥异的音乐。并非所有人都喜欢同一种音乐，而探究人们音乐偏好背后的原因，也并非科学领域的新问题。心理学和营销学这两个研究领域都对此进行了探讨。心理学研究了音乐偏好与心理特征或生活方式之间的关系 [1–7]。但据我们所知，除了唯一一项例外研究 [8]，此前尚未有人考察个人价值观。个人价值观是一种心理特征，体现了人们生活中最看重的事物。营销领域早已认识到，个人价值观有助于预测消费者会对哪些产品属性产生兴趣。随着新技术的发展，了解人们为何喜欢自己喜欢的音乐变得愈发引人关注。流媒体时代几乎让所有人都能接触到各种各样的歌曲，无须依赖实体音乐商店和广播电台等经过人工整理的音乐收藏。例如，深入了解音乐品味有助于改进相关技术，帮助用户浏览日益庞大的音乐收藏。 [8] 采用了一种传统的营销技术，通过访谈将一组参与者的音乐偏好与价值观联系起来。他们收集了一些定性见解，并绘制了一张图谱。然而，尽管这些定性见解颇具参考价值，却无法推广至更广泛的人群。为延续这一研究方向，我们将基于这项先前研究开展一项调查问卷研究，并从略有不同的角度出发，扩大参与者群体，以更好地了解个人价值观与音乐偏好之间的相互作用。

1.1 音乐偏好与心理特征

心理学领域开展了大量研究，试图将音乐偏好与广泛的心理特征联系起来。 [4] 总结了其中大部分研究发现，涉及政治倾向、宗教信仰、性别、族裔、年龄、教育，以及同伴和父母带来的社会影响。人们在青少年时期接触到什么音乐，确实会强烈影响其音乐偏好 [5]。一项大规模调查问卷也从生活方式的角度考察了音乐偏好。 [1–3]

与音乐偏好相关、受到最广泛研究的心理特征之一是人格 [6]。然而，近期的一项荟萃分析 [7]得出的结论是，这些效应相当微弱，因此“人格特质几乎无法解释音乐偏好中的个体间差异”。尽管人格已得到广泛研究，但据我们所知，几乎没有研究探讨个人价值观与音乐偏好之间可能存在的关系。

1.2 个人价值观

个人价值观是身份认同的核心组成部分，体现了人们希望成为怎样的人，也指明了他们在生活中最看重什么。人们会受到内在动机驱动，去消费能够彰显自身个人价值观的产品，甚至不消费会贬损这些价值观的产品。心理学界数十年来一直在研究个人价值观。由于价值观是主观构念，人们建立了许多不同的模型。其中最著名的模型之一是施瓦茨模型 [9], ，该模型本身也历经数次修订。本文采用的版本称为肖像价值观问卷（PVQ） [10,11]。该问卷包含 11 种价值观，分属 4 个类别：保守（关心生活各方面的安全）、开放变革（重视独立与探索）、自我超越（关心世界）和自我提升（关心自身）。这些价值观及其类别依据动机相似性进行排列。

1.3 手段—目的理论

心理学并不是唯一关注个人价值观的领域。营销领域发现，许多消费者选择都受到其个人价值观的驱动：例如，是选择最便宜的选项、知名品牌，还是环保产品。甚至葡萄酒冷却瓶的尺寸或材质，也可能成为基于价值观的偏好的影响因素 [12]。这基于手段—目的理论 [13]，该理论认为，产品的特征是消费者实现某种目的的手段，而这一目的就是他们的个人价值观。

据我们所知，迄今为止，这些关联主要是通过一种称为阶梯访谈法 [12,14] 的营销访谈技术，以更为手工化和定性的方式建立起来的。在访谈过程中，访谈者首先从参与者那里获取一些有关待评估产品的具体特征，例如以玻璃瓶装汽水或含糖量较低。这些元素被称为属性。随后，访谈者会通过询问“这对你来说为什么重要？”或“为什么这对你很重要？”等问题，逐步提升抽象层次，直至参与者的价值观。属性与价值观之间的中间步骤被称为后果。它可以被视为消费者试图通过产品属性获得的效应，例如显得更加老练或提高生产效率。

1994 [15] 的一篇文献综述声称，这一后果的中间步骤是必要的，因为在更大的量表上，它与属性和价值观都有很强的联系，

而属性和价值观彼此之间的联系却很弱。[1]

例如，这项技术曾用于探究人们对智能手机 [16], 住房 [17] 和道德服装 [18]的偏好与价值观之间的关联。

1.4 手段—目的理论与音乐偏好

另一项以音乐为主题的访谈研究分享了一些有价值的见解，探讨个人价值观与音乐之间的关系。本研究旨在从两个主要方面拓展这项研究。首先，通过询问参与者对他人音乐偏好的看法，而非他们自己的音乐偏好，从不同角度考察价值观与音乐偏好之间的关系。这种做法将补充首项研究的视角：人们如何看待自身价值观与偏好之间的关系。

其次，采用更定量的方法，重点运用数据模拟来强化用户实验。这将使这项探索性研究融合定性与定量方法。因此，研究结果将具有更强的普遍性，但并非完全如此。

1.5 研究问题

本研究旨在通过探讨两个主要研究问题，以更定量的方式审视先前一项定性研究的发现。

- 研究问题 1：价值观、后果与音乐属性之间有何联系？（探索性）

  虽然可供参考的先前研究很少，但我们确实预期至少有些价值观会与某些音乐属性相关。为了评估哪些联系对未来研究和应用具有意义，我们需要针对每一种潜在联系回答以下两个问题。

  – 研究问题 1.1：这种联系有多强?
    价值观与音乐属性或后果之间的联系必须足够强，才具有实用价值。确实，如果属性“响亮”仅能很轻微地支持价值观“抱负”，那么这或许并不是值得进一步探究的联系，在应用层面上也未必有多大意义。联系的强度或程度将通过参与者给出的平均分来衡量。平均分与 0 相距越远，联系就越强。该平均分还将反映联系的方向：它是促进还是削弱该价值观。

  – 研究问题 1.2：这种链接有多稳定?
    价值观与音乐属性或后果之间的链接也需要在总体中保持稳定。换句话说，大多数人必须认同这种链接是否存在及其强度。具体而言，在本实验中，我们将通过考察每种可能链接的评分标准差来衡量稳定性。标准差越小，链接就越稳定。

  由于此前缺乏关于这一主题的研究，我们无法对这些链接的具体情况提出明确假设。本研究的主要目标是初步了解哪些链接在不同参与者之间较为稳定，哪些则不太稳定。同时，通过确定值得进一步调查的链接，为这一主题的深入研究提供方向。

[1]遗憾的是，我们未能找到这项主张所依据的研究。

- 研究问题2：我们能否绕过后果步骤，直接将个人价值观与歌曲属性联系起来?

  根据 [15], ，抽象个人价值观与具体产品属性通常需要通过一个中间抽象层级联系起来。然而，最初提出这一主张的研究无法找到。因此，我们希望在自身研究情境中验证这一主张。采用这一额外层级会让参与者的流程更加繁琐，也不够直观。考虑到我们的方法在未来的研究与应用，弄清这一步是否确有必要，将是一项有价值的贡献。

  我们将通过比较两种实验条件的结果来回答这个问题：一种是省去这一中间步骤的直接条件，另一种是包含这一中间步骤的间接条件。

1.6 潜在应用

更深入地了解个人价值观与音乐属性之间的联系，将为推荐系统领域带来有趣的应用。推荐系统是一种旨在帮助用户浏览海量且往往不断扩大的项目集合的算法。传统上，这些系统依赖用户过去在系统中的行为记录：他们消费过什么？喜欢什么？两种主要方法是协同过滤和基于内容的方法。协同过滤方法利用用户与项目之间以往的交互，计算用户之间或项目之间的相似度。随后，它会向当前用户推荐消费偏好相似的用户喜欢的内容，或与当前用户先前喜欢的项目相似的项目。基于内容的方法则利用有关项目本身的显式信息。例如，对于一首歌曲，这些信息可以包括艺术家、发行年份、音乐类型、国家等。随后，它会向用户推荐具有与其此前喜欢或消费过的项目相似特征的项目 [19]。

因此，音乐推荐系统通常会关注用户此前听过的内容，主要将他们引导至相似的歌曲。该领域的部分研究致力于改进推荐系统，将用户引导至更出人意料但仍然优质的推荐内容。我们认为，在推荐系统中使用个人价值观，将有助于提供拓展品味的推荐，同时仍能针对每位特定用户进行定制。事实上，这类推荐系统将不会（或至少较少）依赖用户已经听过的内容。此外，由于个人价值观代表了人们希望成为的样子，用户可能会受到内在动机驱动，去遵循基于个人价值观的推荐。

我们的研究方向旨在通过考察个人价值观与音乐属性之间联系的强度和稳定性来探索这一问题。我们的结果将为了解个人价值观在音乐推荐背景下的潜在用途提供洞见。

2 方法

本研究由多个步骤组成。首先，我们开展了一项简短的先导研究，以测试调查问卷。随后，我们开展了一项大规模模拟研究，既为分析做准备，也估计所需的参与者人数。最后，我们计划通过 Prolific（一个专为科学研究设计的英国众包平台），向一般人群的样本发放调查问卷。

2.1 试点研究

在运行模拟和实验之前，我们进行了一项先导研究。我们的目标是对调查问卷进行预试，确保说明易于理解，

并且问卷不会太长。为此，我们在问卷末尾添加了几个问题，以及一个用于收集反馈的开放式问题。我们还想了解人们平均需要多长时间才能完成问卷，以便为参与者确定适当的报酬。我们的第二个目标是收集一些估算值，以确定运行模拟所需的参数。

事实上，模拟需要估算总体参数才能正常运行。这些参数可以根据以往类似研究或通过先导研究来估算。由于我们找不到与本主题相关的既往研究，因此只能开展先导研究。我们仅从这项先导研究中提取了模拟所需的信息：每个链接评分的标准差、每位参与者评分的标准差，以及两种实验条件下评分稳定性（标准差）差异的平均值。

先导研究邀请了 2 位未参与本项目的同事参加。随后，我们通过 Prolific 招募了 8 位参与者，使样本总数达到 10 人。由于参与者被随机分配到两种实验条件之一，6 人参加了间接条件，4 人参加了直接条件。两种实验条件的完成时间中位数不同，分别为 51:05 分钟和 16:45 分钟。基于这些情况，并考虑到 Prolific 的奖励机制，开展本研究的最佳方式似乎是先进行一项预筛选研究，筛选出不介意参加任一实验条件的参与者。然后，使用他们的 Prolific ID 将其随机分配到任一实验条件，并在 Prolific 平台上分别开展两项独立研究，以便公平地支付参与者的报酬。我们将根据参与者所分配实验条件的完成时间中位数，并按每小时 £9 的标准向他们支付报酬 [2]。

分析 [3] 显示，每个链接评分的平均标准差为 0.41，每位参与者评分的标准差为 0.49。

2.2 模拟研究

在试点研究之后并结合其结果，我们构建了一项模拟研究。模拟研究有几个优点。首先，它们可以让我们预先编写分析脚本，从而在发放调查问卷之前提前发现需要进行的更改。第二个优点是，可以通过多次运行模拟并将结果与预先设定的效应和总体参数进行比较，估算所需的参与者人数。第三个优点是，它还可以用于检验计划采用的分析方法，并发现实际应用中可能并不适当的测量指标和统计检验。当尚无成熟、简便的方法可根据收集的数据回答研究问题时，这一点尤为重要；本研究正是如此。

模拟的代码使用R编写，并基于一份教程 [20]。[4] 总之，该代码返回一组由模拟参与者给出的模拟评分，并将总体参数和效应参数作为输入：

- n_subj：要模拟的参与者总数
- β_0：所有“真实值”评分的总体均值。
- τ_0：被试间随机截距的标准差。这模拟了参与者之间评分行为的差异。例如，有些人会用到量表的全部长度，而有些人几乎从不使用更极端的取值。

[2]这是 Prolific 建议的合理时薪。[3]请参阅随本文附上的代码中的“Pilot Analysis.Rmd”文件。[4]详细且附有注释的代码笔记本可在此处找到：`validation-values-music-map` https://github.com/smamolias/ 这些笔记本也随本文一并提供。

- ω_0：项目间随机截距的标准差。这模拟了不同链接之间的差异。对于大多数参与者来说，有些链接可能比其他链接更明显，因此其标准差会小于其他链接。此参数考虑了这种效应。

- X：实验条件之间评分的原始平均差异。此参数用于模拟实验条件效应导致的评分差异。例如，评分者在间接条件下给出的评分比在直接条件下更极端。

- σ：残差误差的标准差。由于人的情况复杂，并非所有因素都能得到控制，因此总会留下一小部分无法解释的变异。此参数用于模拟这种变异。

根据这些参数，脚本会模拟项目（链接）以及指定数量的参与者。脚本将前一半参与者分配到第一个实验条件（直接条件），后一半分配到第二个实验条件（间接条件）。随后，根据以下方程计算每位模拟参与者在其所属实验条件下对每个项目（链接）给出的评分：

$$score = \beta_0 + \tau_{0s} + O_{0i} + X_i + e_{si}$$

其中，O_{0i} 是项目i的随机截距，X_i 是实验条件对项目i 的效应，e_{si} 是涉及受试者s 和项目i的试次的残差误差。

由于缺乏可供参考的类似既有研究，我们默认假设评分服从正态分布。收集到足够的数据后，我们会检验这一假设是否成立。

2.3 模拟分析

我们对模拟结果进行了不同的分析。首先，我们计算了每次模拟的 Cohen d效应量，以获得标准化的、因而更易解释和比较的效应量。

第二项与教程中的相同：进行功效分析，以了解功效如何根据条件效应和误差而变化。为此，我们使用以下方程运行线性模型：

$$score = condition + (1|participant_i d) + (1|item_i d)$$

并提取 p值。然后，对于每批 1100 次模拟，我们只需计算低于显著性阈值的 p值所占的比例。为提高灵活性，我们将 0.05 和 0.01 都作为显著性水平。

我们还对缩放后的得分执行了相同的步骤。在实际实验中，参与者只能给出 -1 到 1 之间的评分，但模拟在生成评分时没有相同的边界。我们使用 S形函数对此进行了校正。我们同时保留了缩放后的得分和未缩放的得分，因为二者各有优缺点。

第三项分析使用两个单侧t检验，并辅以等效性检验（最小效应量为0.2），比较两种实验条件下每个链接评分均值的原始分布。我们同时检验差异性和等效性，以区分某个效应是显著的、过小而不具有研究意义/不存在，还是处于无法解释的灰色地带。根据Cohen d的解释指南 [21]，0.2是可忽略效应与小效应之间建议采用的阈值，因此我们将0.2选为具有研究意义的最小效应量。

2.4 模拟迭代过程

我们使用以下参数运行了 1100 次模拟：

- $\beta_0 = 0$
- $\tau_0 = 0.486$
- $\omega_0 = 0.410$
- 受试者人数 $= 100$, 200, 400, 600, 800, 1000
- 条件效应 $= 0$, 0.024,0.05,0.1, 0.5, 0.8
- $\sigma = 0.1$、0.2、0.5、0.8 和 1

参数 τ_0 和 ω_0 （以及数值为 0.024 时的条件效应）基于预试验结果，而 σ 是一个自由度，我们将对其测试多个取值。

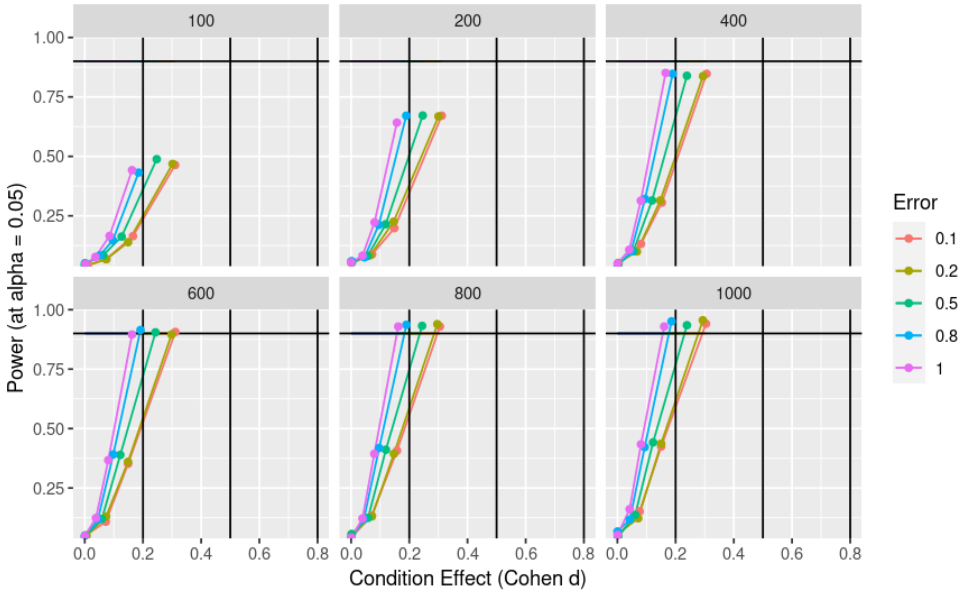
我们对每次模拟的结果采用线性回归模型，以收集计算每批1100次模拟总体检验效能所需的p值。所使用的方程如下：

$$link\text{-}rating = \beta_0 + (1|participants) + (1|items)$$

其中，参与者和项目都被视为随机效应。事实上，我们只有总体中的一个样本，而且每位参与者都与其他参与者存在许多差异，包括他们使用调查问卷中量表的方式。并且，元素（属性与后果）的列表，因此也包括本实验中使用的配对，未必是穷尽的。此外，有些配对可能比其他配对更容易评分。

2.5 模拟结果

图1. 模拟结果的Morey图。这些图展示了检验效能随参与者数量、误差和条件效应变化的情况。每张图代表不同的效应量。这一次，条件效应并非原始值，而是转换为 Cohen d。水平线表示90%检验效能，垂直线表示不同效应量解释（可忽略、小、中等、大）之间的传统截断值。



我们从这些结果中寻求的主要信息，是估计需要多少参与者才能达到合理的检验效能。我们选择给出

我们为实验设定了90%检验效能，并决定不关注可忽略的效应。因此，我们希望数据点位于90%竖线的上方，并处于0.2横线的右侧。基于此，我们选择招募600名参与者作为样本。

2.6 总体

我们将通过 Prolific 招募参与者。根据模拟结果，我们将招募 600 名参与者，仅设定两项条件：年龄须满 18 岁（出于伦理原因），且须能流利使用英语（以便充分理解调查问卷）。参与者完成调查问卷后，将按每小时 9 英镑的标准获得报酬 [5]。根据试测中对各实验条件完成时间的估算，本研究的费用约为 3050 英镑 [6]。

2.7 伦理

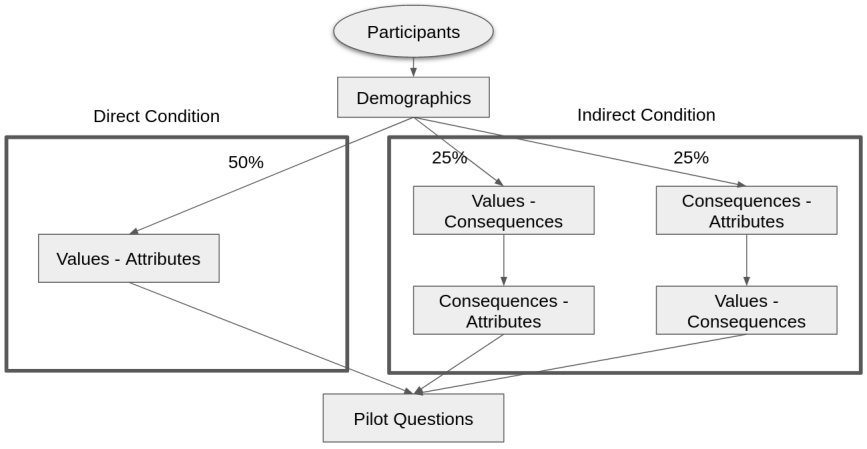
本实验已获代尔夫特理工大学伦理委员会批准。我们的数据管理计划也已获批准。简而言之，参与者将被告知其个人信息和回答将如何处理。参与实验前，他们必须表示同意。为尊重参与者的隐私，人口统计信息将以汇总形式报告，且不会在研究团队之外共享。参与者对链接的评分将在线公开，并收录在随本研究完整版一并提供的数据集中。只有匿名参与者编号会与评分相关联。

3 材料

3.1 调查问卷

该调查问卷基于一项先前的访谈研究的结果构建而成 [8]。调查问卷采用了在这些访谈中 elicited 的属性与后果，但排除了被认为与本研究情境无关的内容（因为信息量不足），例如“音乐类型”或“音乐特征” [7]。

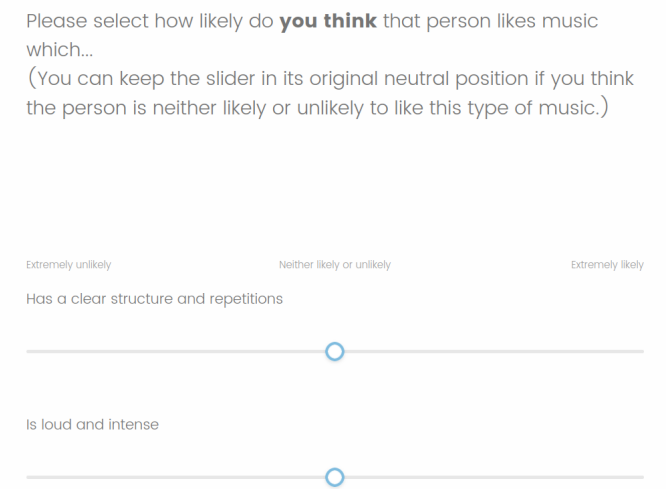
图 2. 调查流程



[5]这相当于 10.32 欧元或 10.4 美元 [6]这相当于 3470.44 欧元或 3588.48 美元 [7]本研究的完整关键词列表见补充材料。

该调查问卷由一个标准的人口统计学部分开头组成，其中我们收集参与者的年龄、申报的性别以及他们成长所在的国家。这些问题的唯一目的，是评估样本的多样性。我们特别询问参与者成长所在的国家，而不是他们的国籍或原籍国，因为人们在青少年时期接触何种音乐，会对其音乐偏好的形成产生很大影响（见第 1.1 节）。我们还询问参与者能够流利使用哪些语言，以核实英语是其中之一。

图 3. 调查问卷中的量表示例



随后，调查问卷分为两个实验条件（以回答研究问题 2），如图2所示。在第一个实验条件下，参与者需要将个人价值观与音乐属性直接关联起来。价值观通过价值观肖像呈现，即对某个假想人物价值观的描述。这些肖像由施瓦茨肖像价值观问卷（PVQ）[10, 11], 中的项目组合而成，仅做了少量修改，以统一句式并使文本更流畅。没有修改任何关键词 [8]。针对每幅肖像，都会提出以下问题：“你认为这个人喜欢以下类型音乐的可能性有多大.....”问题下方列出一系列没有显示数值的连续量表，范围从“极不可能”到“极有可能”，中间点和起始点为“既不可能也不不可能”，如图3所示。每个量表对应 10 种属性中的一种。我们选择使用连续量表，以便更精细地收集和分析数据 [22, 23]。我们采用中性的中间点，因为我们不认为每种属性都与每种价值观相关。这个中间点降低了参与者的认知负荷：如果他们认为这种关联不存在，就可以直接忽略相应的量表／问题。这样，他们需要评分的量表就更少；同时，我们也能明确得知参与者认为这一特定的属性—价值观关联并不相关。

第二种条件包含两组问题，按照随机顺序呈现，如图2右侧所示。其中一组要求参与者将价值观与后果联系起来，方式与第一种条件下参与者将价值观与属性联系起来的方式相同。另一组问题旨在要求参与者将后果与属性联系起来。参与者会被问道：“您在多大程度上同意以下陈述？具有 [插入后果] 的音乐同时也是：”。随后是一系列没有数值、标注为“极度不同意”至“极度同意”的连续量表，其中间点和起始点为“既不同意也不反对／不适用”。量表的格式及其背后的依据

[8] 每幅肖像的完整文本见补充材料。

与其他问题相同。只有标签略作修改，以便与问题的措辞在语法上保持一致。使用相同的量表将有助于分析，尤其能够对各问题进行比较。这一优势足以抵消使第二种条件比简单的二元回答格式更长、更繁琐这一缺点，而简单的二元回答格式正是我们最初的设想。

在两种实验条件下，价值观、属性与后果的呈现顺序均为随机。第一种实验条件下，每位参与者都会看到所有可能的价值观—属性配对；第二种实验条件下，每位参与者都会看到所有可能的价值观—后果配对以及后果—属性配对。这样的随机化有助于控制项目呈现顺序可能造成的任何非预期效应。

4 分析计划

分析所用的代码以 R Markdown 编写，文件名为 Analysis.Rmd，力求尽可能易于理解。相关函数保存在单独的 Functions.Rmd 文件中。[9]

4.1 预处理

首先，我们将从实验结果中排除未完成的调查问卷，以及在人口统计部分的任何语言问题中均未填写英语的参与者。

在所有调查问卷中，我们预计参与者会将一定数量的链接视为不存在或无关。我们在每道题的说明中要求参与者在这种情况下将光标留在量表的起始位置。但 Qualtrics 会将这些情况记为缺失值，而在我们的语境中，它们实际上应为 0（“既不可能也不不可能”）。因此，每位参与者在其所分配实验条件下的所有缺失值评分都将改为 0。

4.2 人口统计学

我们只会将人口统计学问题用于评估样本的多样性。因此，我们只会计算并报告参与者的年龄均值和标准差，以及参与者在性别和成长国家方面的分布。我们不会探究这些变量对调查问卷答案可能产生的任何效应，因为这与我们的研究问题相距甚远，而且需要采用不同的研究设计开展另一项研究。

4.3 研究问题1：个人价值观与音乐偏好之间存在什么联系?

为了回答这个问题，我们将从两个方面考察参与者对这些联系的评分：**效应强度**和**评分者一致性**。

4.3.1 研究问题1.1：这些联系有多强?

前者将揭示效应有多强，也因此有多值得关注。为此，我们将计算每个链接的评分均值。该均值越远离 0，效应就越强；越接近 1，效应就是正向的：例如，

[9]代码已随本文一并提供。

音乐属性 B 会强化价值 A。如果评分均值更接近 -1，那么音乐属性 B 就会削弱价值 A。

4.3.2 研究问题 1.2：这些链接有多稳定?

后者将有助于了解效应有多稳定。大多数参与者是否倾向于认同链接是否存在、其方向和强度？关于评分者一致性的测量方法有很多，但对于在特定情境下应采用哪一种，人们鲜有共识。大多数方法是针对项目整体计算的，但要回答我们的研究问题，我们需要一种能够针对每个链接（项目）单独计算的一致性测量方法。因此，我们决定采用每个链接评分的标准差。我们认为这是最合适的方法，因为它能非常清晰地反映评分围绕均值的离散程度。

从略带探索性的角度出发，我们还将计算组内相关系数（组内相关系数，ICC），以获得更全面的概况。组内相关系数是一项用于计算所有项目整体评分者一致性的指标。它将有助于了解总体而言，人们是否倾向于认同音乐偏好与个人价值观之间的链接。

由于缺乏可供参考的既有研究，目前没有明确依据可以确定应使用何种截断值，以判断某个链接是否足够强、值得关注，以及一致性是否足够高，从而认为该链接的这种表征在我们的样本中是稳定的。因此，我们将对这些链接进行排序，并以更具定性的方式解读该分析的结果，以进一步了解哪些链接相比其他链接最强、最稳定。这些链接将是后续研究中最值得进一步考察的对象。

4.4 研究问题 2：我们能否直接将个人价值观与歌曲属性联系起来?

本实验的两个实验条件旨在回答这一研究问题。我们希望了解，在我们的音乐偏好语境中，正如 [15], 所声称的那样，是否需要一个中间抽象步骤来连接具体的属性和价值观。为评估这一点，我们将使用绝对评分进行两次单侧 t 检验和一次等效性检验，比较参与者在两个实验条件下的绝对评分。这些检验将初步揭示两个实验条件之间绝对评分分布的潜在差异。等效性检验将用于衡量该效应是否小到不足以引起关注。

致谢

我们谨向所有在本项目过程中给予帮助的人表示感谢。特别感谢 Nava Tintarev 教授在项目初期提供的宝贵意见，以及 Chelsea Parlett Pelleriti 博士对本研究统计部分提出的极具价值的见解。

补充材料

4.5 条目列表

4.5.1 属性

- 有趣
- 带来活力

- 让他们表达自己的身份认同
- 引发负面情绪
- 支持创造力
- 令人熟悉，并唤起旧日回忆和怀旧之情
- 独一无二
- 让他们感到强大
- 让他们有所感受
- 具有挑战性，而且不枯燥
- 让人发现新事物
- 具有意义和真实性
- 联结人们
- 帮助他们集中注意力，提高效率
- 展现艺术家的努力、才华和技能
- 帮助他们完善自我
- 令人放松

4.5.2 后果

- 表达一种情感
- 是原创的
- 和谐且旋律优美
- 简单且易于理解
- 有一个故事
- 有清晰的结构和重复
- 响亮而强烈
- 即兴且自发
- 有柔和的声音

4.5.3 肖像

- 从众主义：“他们希望避免做任何会被人们认为是错误的事情，并相信人们应该始终听从吩咐、遵守规则，即使没有人在旁监督也是如此。他们尽量不去打扰或惹恼他人。对他们而言，始终举止得体、顺从并礼貌地对待他人，以及尊重父母和长辈，都非常重要。”

• 传统：“对他们而言，遵守所学到的习俗很重要，他们认为最好以传统方式做事。宗教信仰对他们很重要，他们努力遵循宗教的要求。他们认为不应索取超出自己所有的东西，而人们应对此感到满足。保持谦逊和谦虚对他们而言也很重要，他们尽量不让自己受到关注。”

• 仁爱：“他们希望投身于关爱亲近的人，认为帮助身边的人非常重要，也希望关心他们的福祉。对他们而言，忠于朋友并回应他人的需求也很重要，他们会尽力支持熟识的人。宽恕伤害过自己的人对他们来说也很重要。他们努力发现他人身上的优点，不记仇。”

• 普世主义：“他们相信世界各地的人都应和谐共处，并认为促进世界各群体之间的和平十分重要。他们希望每个人都得到公正对待，即使是素不相识的人也不例外。他们认为，世界上的每个人都应受到平等对待，社会中的弱势群体也应得到保护，这一点十分重要。他们相信每个人在人生中都应拥有平等的机会。他们努力理解他人，即使与对方意见不合也不例外。他们认为，倾听与自己不同的人十分重要。关爱环境对他们来说很重要，他们也希望顺应自然、与自然和谐相处。他们坚信人们不应改变自然，而应爱护自然。”

• 自我导向：”独立自主对他们来说很重要。他们喜欢用自己原创的方式做事。对他们而言，想出新点子并发挥创造力，以及自行决定要做什么，都很重要。他们喜欢依靠自己，也喜欢自由地为自己规划和选择活动。他们还喜欢保持好奇，尝试了解各种各样的事物，并认为对事物保持兴趣很重要。”

• 刺激：”拥有令人兴奋的生活对他们来说很重要。他们喜欢惊喜，总是在寻找新鲜事物来尝试。他们总是寻求冒险，也喜欢冒险行事。他们认为在生活中体验许多不同的事情很重要。”

• 享乐主义：”他们非常希望享受生活。尽情享受生活中的乐趣对他们来说很重要。他们喜欢“宠爱”自己，并寻找一切可以享乐的机会。做能带给自己快乐的事情对他们来说很重要。”

• 成就：“对他们而言，展现自己的能力很重要，他们认为有抱负也很重要。取得巨大成功、在人生中出人头地对他们来说都很重要。他们努力做得比别人更好，也喜欢给人留下深刻印象。他们想展示自己的能力有多强，并希望别人欣赏他们的所作所为。”

• 权力：“他们喜欢当领导者，总想成为那个做决定的人。对他们来说，掌管一切并告诉别人该做什么很重要：他们希望别人按照他们说的去做。富有对他们来说也很重要：他们想拥有很多金钱和昂贵的物品。”

• 安全：“生活在安全的国家和安全的环境中对他们来说很重要。他们会避开任何可能危及自身安全的事物，并认为国家必须时刻防范来自内部和外部的威胁。拥有稳定的政府

对他们来说很重要。他们关心社会秩序是否得到维护。事物井然有序、整洁对他们来说很重要。他们非常不喜欢杂乱无章。他们也努力避免生病，因为保持健康对他们来说同样非常重要。”

参考文献

1. North AC, Hargreaves DJ. 音乐偏好的生活方式相关因素：1. 人际关系、居住安排、信仰与犯罪。《音乐心理学》. 2007;35(1):58–87. 2. North AC, Hargreaves DJ. 音乐偏好的生活方式相关因素：2. 媒体、闲暇时间与音乐。《音乐心理学》. 2007;35(2):179–200. 3. North AC, Hargreaves DJ. 音乐偏好的生活方式相关因素：3. 旅行、金钱、教育、就业与健康。《音乐心理学》. 2007;35(3):473–497. 4. Laplante A. 改进音乐推荐系统：我们能从音乐品味研究中学到什么？载于：ISMIR；2014. 第451–456页. 5. Laplante A. 谁影响青少年的音乐品味？一项关于社交网络中人际影响的研究。载于：第二届 ACM 用户中心与多模态策略音乐信息检索国际研讨会论文集。ACM；2012. 第37–42页. 6. Dunn PG, de Ruyter B, Bouwhuis DG. 深入理解音乐偏好、聆听行为与人格之间的关系。《音乐心理学》. 2012;40(4):411–428. 7. Sch¨afer T, Mehlhorn C. 人格特质能否预测音乐风格偏好？一项荟萃分析。《人格与个体差异》. 2017;116:265–273. 8. Manolios S, Hanjalic A, Liem CCS. 个人价值观对音乐品味的影响：迈向基于价值观的音乐推荐。载于：第13届 ACM 推荐系统大会论文集；2019. 第501–505页. 9. Schwartz SH. 基本人类价值观：概述。摘自 http://www yourmorals org/schwartz. 2006;. 10. Schwartz S, Melech G, Lehmann A, Burgess S, Harris M, Owens V. 采用不同测量方法拓展基本人类价值观理论的跨文化效度。《跨文化心理学杂志》- J CROSS-CULT PSYCHOL. 2001;32:519–542. doi:10.1177/0022022101032005001. 11. Schwartz S. 载于：《个体价值普遍性理论的稳健性与成果丰硕性》；2005. 第56–95页. 12. Reynolds TJ, Gutman J. 阶梯理论、方法、分析与解释。《广告研究杂志》. 1988;28(1):11–31. 13. Gutman J. 基于消费者分类过程的手段—目的链模型。《营销学杂志》. 1982;46(2):60–72. 14. Saaka A, Sidon C, Blake BF. 阶梯法：操作指南——兼谈注意事项。《消费者行为研究报告：操作指南系列》。俄亥俄州：克利夫兰州立大学。2004;.

15. Van Raaij WF, Verhallen TM. 特定领域市场细分。《欧洲市场营销杂志》. 1994;. 16. Leitner M, Wolkerstorfer P, Sefelin R, Tscheligi M. 移动多媒体：运用手段—目的理论识别用户价值观。载于：第十届移动设备与服务人机交互国际会议论文集；2008. 第167–175页. 17. Collen H, Hoekstra J. 价值观作为住房属性偏好的决定因素。《住房与建成环境杂志》. 2001;16(3-4):285–306. 18. J¨agel T, Keeling K, Reppel A, Gruber T. 个人价值观与伦理服装消费中的动机复杂性：一种手段—目的方法。《市场营销管理杂志》. 2012;28(3-4):373–396. 19. Ricci F, Rokach L, Shapira B. 载于：Ricci F, Rokach L, Shapira B，编者。《推荐系统：导论与挑战》。美国马萨诸塞州波士顿：Springer US；2015. 第1–34页. 获取地址：https://doi.org/10.1007/978-1-4899-7637-6_1. 20. DeBruine LM, Barr DJ. 通过数据模拟理解混合效应模型。《心理科学方法与实践进展》. 2021;4(1):2515245920965119. 21. Cohen J. 心理学定量方法：统计功效入门。载于：《心理学通报》。Citeseer；1992. 22. Toepoel V, Funke F. 滑块、视觉模拟量表还是按钮：格式与量表对移动端和桌面端调查的影响。《数学人口学研究》. 2018;25(2):112–122. 23. Chyung SY, Swanson I, Roberts K, Hankinson A. 循证调查设计：连续评分量表在调查问卷中的应用。《绩效改进》. 2018;57(5):38–48.