

《心理学报》审稿意见与作者回应

题目：分析思维降低情感预测影响偏差

作者：耿晓伟 刘丹 牛燕华

第一轮

审稿人 1 意见：

意见 1：本研究关注的问题，“如何降低情感预测偏差”是一个很有意思的问题，既有理论价值，又有实践意义。本研究问题意识明确，通过引入分析思维来回答这一问题，通过一个预实验和两个正式实验收集数据检验理论的合理性，有重要的创新，对决策心理学有重要的贡献。然而，文章还有诸多问题，留待作者认真思考、修改。

首先，理论。文章要解决的问题是“文章聚焦的问题应是“如何降低情感预测偏差”。文章依据的理论是双系统理论。但是双系统理论，虽被称作理论，但可能不是一个心理学意义上的中观层次的理论。或许它可以被称作一种思想，一种取向，或一个原理(principle)。因为双系统理论没有明确的适用范围，且只能产生原则性预测，而不能产生具体的预测。这也是为什么本研究不可太可能在理论上发展，而只能验证双系统理论的原因（如文章总讨论 5.3 第二段末句所写“.....[本文]为双系统理论提供了进一步证据”）。实际上，依我看来，本研究是分析思维在情感预测领域的一种应用。作者在文献综述部分也提到分析思维启动降低个体宗教信仰程度的诸多研究。那么，为什么不把这些研究和理论介绍清楚，借鉴过来，作为自己提出假设的直接基础呢？作者的创新是引入分析思维，但对分析思维却没有定义，也没有介绍之前研究如何操纵和测量分析思维。这是本文应该具体着力之处。同样由于理论基础（双系统理论）太过笼统和宽泛，造成本文的假设：分析思维降低情感预测，进而降低情感预测偏差，无法深入。作为读者的我，非常感兴趣明白这个假设背后的个中道理。为什么分析思维可以降低情感预测偏差呢？这个道理跟之前人发现的道理有何不同，又有何关联？对该问题的回答相信会让本文有质的飞跃和提升。落实到具体的研究上，这种深入的思考不仅可以使引言更加聚焦，让总讨论更加言之有物，也会让本研究的设计更加精细。

回应：非常感谢您的建议。的确如您所说的，双系统理论是本研究的理论基础。尽管目前有很多关于分析思维的研究，这些研究分别考察了分析思维对宗教信仰的影响（Gervais & Norenzayan 2012; Stagnaro, Ross, Pennycook, & Rand, 2019），对偏见的影响（Yilmaz, Karadöller, & Sofuoğlu, 2016; Franks & Scherr, 2017），以及对共谋理论信念的影响（Swami, Voracek, Stieger, Tran, & Furnham, 2014）等，然而这些研究都没有明确给出分析思维定义。在这些研究中，直觉思维即系统 1（直觉系统），分析思维即系统 2（分析系统）。在本研究中，我们沿用了前人研究的做法，认为分析思维即系统 2。

就分析思维的测量和操纵而言，目前分析思维的测量工具一般采用 Cognitive Reflection Task (CRT)（Gervais & Norenzayan 2012; Frederick, 2005; Stagnaro, et al., 2019）。在 CRT 任务中直觉做出的答案是错误的，只有当分析思维压倒直觉思维的时候，答案才会是正确的，因此 CRT 的成绩可以用来代表分析思维。目前分析思维的操纵方法有几种，包括图片启动（Gervais & Norenzayan, 2012; Franks & Scherr, 2017）、语言流畅性任务或混词组句启动（Gervais & Norenzayan 2012; Swami et al., 2014）等。我们已经在文献综述部分补充了分析思维的定义以及相关理论，详见 1 引言部分的高亮部分。

关于“为什么分析思维会减低情感预测偏差”。情感预测本质上属于一种对情绪的自陈报

告。根据情绪自陈报告的可获得性模型(accessibility model of emotional self-report) (Robinson & Clore, 2002), 人们对情绪进行自陈报告时遵循相对可获得性(relative accessibility)原则, 也就是说, 人们根据那些可获得性高的信息做出情绪自陈报告。越具体的信息, 相对可得性越高(Robinson & Clore, 2002)。当人们对未来事件带来的情感进行预测时, 由于无法获得任何具体的情景记忆信息, 此时关于未来事件的信念信息具有更高的可获得性, 因此, 情感预测直觉地基于人们已有的信念。根据双系统理论观点, 分析思维会削弱或者压倒直觉思维 (Evans, 2008)。已有研究表明, 启动分析思维会削弱人们的宗教信仰等各种信念。因此, 我们推测分析思维也会削弱人们关于未来事件的信念, 从而降低人们对未来事件情感预测的强度, 降低影响偏差。

意见 2: 其次, 方法。2.2 第二段, 作者选取认知反应测试作为操纵检验的方法, 但以前研究定义的分析思维跟本研究定义的分析思维是一致的吗? 如果是一致的, 那么为什么不直接借鉴以往研究的分析思维的启动方法呢? 建议作者对分析思维的概念和操作定义都做适当界定, 并厘清与相关概念(比如计算思维)的联系和区别。实验 1 选取的三种情绪词(高兴、开心、快乐), 效价都是正的。作者应说明为什么选这三个情绪词, 以及为什么没有选效价为负的情绪词。尤其是当被试预期自己被拒绝时, 拿正效价情绪词来提问就比较奇怪了。实验 1 的 3.3 实验程序部分, 需要进一步说明。3.3 提到, “要求被试预测他的分配方案被接受或被拒绝后自己的情感”而后面又提到, “最后, 实验者给被试随机反馈为接受或者拒绝”接受/拒绝条件, 不知道是如何操纵的。前面被试被要求预测接受或拒绝后的情感, 但后面拒绝或接受的反馈又是随机给出的。那么, 如果前面被试被要求预测自己被接受后的情感, 而后面实验者给被试随机反馈为拒绝的话, 被试的情感预测偏差怎么计算呢? 不知道是不是前面被试需要预测的是接受&拒绝两种条件下自己的情感呢? 作者能否将实验流程写的更清楚易懂一些? 3.4.1, “接受或拒绝方案主效应显著”。但结合后面的图 1, 接受或拒绝方案的主效应显著, 是否是两种条件下情感预测偏差的效价不同所致? 可否直接用两种条件下, 情感预测偏差的绝对值, 进行计算? 进一步, 既然接受/拒绝条件的主效应显著, 那么能否分析一下这个显著的主效应说明什么问题? 4.2.2 中因变量的测量为“自编情感评定量表”, 设定积极情感与消极情感, 分别有几个项目? 共几个项目? 积极、消极情感词是如何选定的, 是否可以做一个简单说明? 4.5 提到“生育孩子对绝大多数人来说都是一个积极生活事件”, 是一个论点, 未给出相关论据。简单一点的话, 是否可以通过给出研究二中积极和消极情感的平均数来支持自己的观点呢? 数据在表 4 里面有。从数据看是支持作者观点的。作者可以搬运一下。

回应: (1) 2.2 第二段, 作者选取认知反应测试作为操纵检验的方法, 但以前研究定义的分析思维跟本研究定义的分析思维是一致的吗? 如果是一致的, 那么为什么不直接借鉴以往研究的分析思维的启动方法呢? 建议作者对分析思维的概念和操作定义都做适当界定, 并厘清与相关概念(比如计算思维)的联系和区别。

回答: 感谢您的建议。本文关于分析思维的定义与前人研究中的定义是一致的, 分析思维即系统 2。Gervais 和 Norenzayan 2012 年在 Science 上发表关于分析思维启动对宗教信仰影响的文章以后, 后续有不少研究者 (Swami, Voracek, Stieger, Tran, & Furnham, 2014; Yilmaz, Karadöller, & Sofuoğlu, 2016; Franks & Scherr, 2017; Stagnaro, Ross, Pennycook, & Rand, 2019) 进行了关于分析思维的实证研究, 这些研究绝大部分都借鉴了 Gervais & Norenzayan (2012) 研究中所采用的分析思维启动方法。本研究采用的思考者图片启动分析思维的方法借鉴了 Gervais & Norenzayan (2012) 和 Franks & Scherr (2017) 的做法。

分析思维与计算思维不同。计算思维是指一种无意的认知倾向, 用数学的方式分析非量化的问题 (an unintended cognitive predisposition to analyze non-quantitative problems

mathematically), 更广泛地说, 是指一种将所有正在考虑的内容、特点等都转换为一个数值以进行质的和数的特点的比较 (reducing “all the relevant features and components under consideration to a single value or utility metric that allows the comparison of many qualitatively and quantitatively diverse features”) (Fiske, 1992; Wang, Zhong, & Murnighan, 2014)。

参考文献:

- Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science*, 336, 493–496.
- Swami, Voracek, Stieger, Tran, & Furnham. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, 133, 572-585.
- Yilmaz, O., Karadöler, D. Z., & Sofuoğlu, G. (2016). Analytic thinking, religion and prejudice: an experimental test of the dual-process model of mind. *The International Journal for the Psychology of Religion*, 26(4), 360–369.
- Franks & Scherr. (2017). Analytic Thinking Reduces Anti-Atheist Bias in Voting Intentions. *The International Journal for the Psychology of Religion*, 27 (3),129-140.
- Stagnaro, M. N, Ross, R. M., Pennycook, G., & Rand, D.G. (2019). Cross-cultural support for a link between analytic thinking and disbelief in God: Evidence from India and the United Kingdom. *Judgment and Decision Making*, 14(2), 179–186.
- Wang, Zhong, & Murnighan. (2014). The social and ethical consequences of a calculative mindset, *Organizational Behavior and Human Decision Processes*, 125, 39-49.

(2) 实验 1 选取的三种情绪词 (高兴、开心、快乐), 效价都是正的。作者应说明为什么选这三个情绪词, 以及为什么没有选效价为负的情绪词。尤其是当被试预期自己被拒绝时, 拿正效价情绪词来提问就比较奇怪了。

回答: 在情感预测的研究中, 不管对积极事件还是消极事件的预测, 研究者通常采用 how happy/pleasant 来提问, 让被试进行情感预测。例如, 在 Gilbert 等人 (1998) 关于情感预测的经典研究中, 让那些没有经历过分手的被试预测分手后的情感 “how happy they would be 2 months after experiencing such a dissolution”, 评分采用 7 点量表, 从 1 (not happy) 到 7 (very happy)。Hoerger 等人 (2016) 的研究中, 让被试对支持的球队赢球或输球后的情感进行预测, “how happy they would feel the following Monday both in the event of a win and a loss”。借鉴前人的做法, 在实验 1 中, 我们要求被试对接受或拒绝后的情感进行预测都采用了同样的问题。尽管问卷中只采用了正效价的情绪词 (高兴、开心、快乐), 然而被试在评价的时候却包含了消极情绪, 采用 5 点量表, 1 代表非常不高兴/不开心/不快乐; 5 代表非常高兴/开心/快乐, 得分越高表示积极情感越强烈, 得分越低则表示消极情感越强烈。另外, 之所以选择高兴、开心、快乐这三种情绪词, 是借鉴前人的研究中常用的 happy/pleasant 进行的汉化。

参考文献:

- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality & Social Psychology*, 75(3), 617–638.
- Michael Hoerger, Ben Chapman and Paul Duberstein. (2016). Realistic affective forecasting: The role of personality. *Cognition and Emotion*, 30(7), 1304-1316.

(3) 实验 1 的 3.3 实验程序部分, 需要进一步说明。3.3 提到, “要求被试预测他的分配方案被接受或被拒绝后自己的情感”而后面又提到, “最后, 实验者给被试随机反馈为接受或者拒绝” 接受/拒绝条件, 不知道是如何操纵的。前面被试被要求预测接受或拒绝后的情感, 但后面拒绝或接受的反馈又是随机给出的。那么, 如果前面被试被要求预测自己被接受后的情感, 而后面实验者给被试随机反馈为拒绝的话, 被试的情感预测偏差怎么计算呢? 不

知道是不是前面被试需要预测的是接受&拒绝两种条件下自己的情感呢？作者能否将实验流程写的更清楚易懂一些？

回答：感谢您的建议。实验流程部分我们确实写的不够清楚。在情感预测时，被试同时预测了接受后的情感和拒绝后的情感，在情感体验时，如果反馈为接受的话，被试则只需要回答接受条件下的情感体验；如果反馈为拒绝的话，被试则只需要回答拒绝条件下的情感体验。计算情感预测偏差时，接受条件下，只采用接受条件的情感预测减去接受条件的情感体验（拒绝条件下的情感预测就没有用了）；拒绝条件下，只采用拒绝条件的情感预测减去拒绝条件的情感体验（接受条件下的情感预测就没有用了）。我们在正文中也进行了补充。

（4）3.4.1，“接受或拒绝方案主效应显著”。但结合后面的图 1，接受或拒绝方案的主效应显著，是否是两种条件下情感预测偏差的效价不同所致？可否直接用两种条件下，情感预测偏差的绝对值，进行计算？进一步，既然接受/拒绝条件的主效应显著，那么能否分析一下这个显著的主效应说明什么问题？

回答：感谢您的建议。接受或者拒绝主效应显著的确是情感预测的效价有关，接受条件下的情感预测（ $M=4.33$, $SD=0.71$ ）显著高于拒绝条件下的情感预测（ $M=2.15$, $SD=0.85$ ）， $t(238) = 21.71$, $p = 0.001$ ，这说明接受条件下被试预期的情感更积极，拒绝条件下被试预期的情感更消极。

（5）4.2.2 中因变量的测量为“自编情感评定量表”，设定积极情感与消极情感，分别有几个项目？共几个项目？积极、消极情感词是如何选定的，是否可以做一个简单说明？

回答：感谢您的建议。自编情感评定量表中，积极情感包含高兴、快乐、幸福，消极情感包含郁闷、疲惫、心烦。这些情感词是借鉴 Wirtz, Kruger, Scollon, & Diener（2003）中用到的积极和消极情感词，从中选取了与生育二孩相关比较密切的情绪词。

参考文献：

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, 14(5), 520–524.

（6）4.5 提到“生育孩子对绝大多数人来说都是一个积极生活事件”，是一个论点，未给出相关论据。简单一点的话，是否可以通过给出研究中积极和消极情感的平均数来支持自己的观点呢？数据在表 4 里面有。从数据看是支持作者观点的。作者可以搬运一下。

回答：非常感谢您的建议。我们已经在文中补充了相应的内容。

意见 3：第三，写作。由于文章聚焦的问题是“如何降低情感预测偏差”，故文章引言部分，1.1 前两段有关情感预测偏差的文献可以一笔带过，而重点介绍以往有哪些研究，有哪些理论，关注了“如何降低情感预测偏差”这个问题的。比如，文章提到的去焦点化训练、适应训练，其理论是怎样的，有哪些具体的研究。引言中对以往研究如何解决“降低情感预测偏差”的介绍，也是本研究理论贡献的一个参照点。总讨论中可以具体展开，本研究发现的分析思维的方法与之前的方法有何不同，机理差异所在，优势又在什么地方。

回应：非常感谢您的建议。我们修改了引言和总讨论部分。

意见 4：最后，细节。2.1 被试部分，预估样本量，应写明按照何种原理预估，而不能让一个免费下载的软件背锅。况且，Gpower 作为术语第一次出现，也应提供一句话的说明及其来源。2.2 第一段最后一句。控制条件只要求被试看四张几何图片吗？不需要被试回答问题吗？那如何知道实验组是看思考者图片引发的分析思维，还是因为被要求自己观察引发的分析思维？这一问题，也同样反映出作者对分析思维的界定还需要思考。预实验中，统计检验力被设定为 0.8，实验 1 的统计检验力被设定为 0.9；预实验中写明了双侧 α ，但实验 1 中就

没有表明是双侧还是单侧 α 。图 1 的背景色，请设置为白色或透明色吧，现在是绿色。

回应：非常感谢您的建议。

(1) 我们在文中补充了 Gpower 的说明和来源。

(2) 实验 1 中为了保证启动效果，保证被试认真观看《思考者》图片，启动组让被试回答两个简单的问题；控制组没有回答问题。预实验发现，观看《思考者》图片后，被试在 CRT 中的得分显著高于控制组，说明启动是有效的。在正式实验中，启动程序与预实验完全相同，因此我们认为启动也是有效的。

(3) 预实验中是被试间 t 检验，因此是双侧 α ，实验 1 中是采用的方差分析，因此是单侧 α ，我们对此在文中进行了补充。另外，我们将预实验和实验 1 的统计检验力都统一为 0.8。

(4) 图 1 的背景色改为透明色。

.....

审稿人 2 意见：

意见 1：《分析思维降低情感预测影响偏差》一文，根据双系统理论以及情绪报告的获得性模型，考察了分析思维对情感预测影响偏差的影响。两个实验设计严谨，结果可靠；就理论意义阐述等方面提出以下建议：

1. 如何用双系统理论来解释情感预测影响偏差 (affective forecasting impact bias, 以下简称 AFIB) 本身的机制，建议作者对这部分理论以及已有研究进行更为清晰的阐述，必要的话可以考虑重新调整综述部分的文章结构；正如去焦点化训练，是根据 AFIB 的 focalism 理论而提出的，作者提出分析思维作为降低偏差的策略，也需要回到 AFIB 本身的过程机制中。这也能为 AFIB 的作用机制提供新的理论与解释思路，提升本研究的理论意义。

回应：非常感谢您的建议。我们已经修改了文献综述部分，详见文中 1 引言部分。

意见 2：实验一结果中，分析思维与两种方案的交互作用显著，如何理解这一交互作用？

回应：非常感谢您的意见。实验 1 结果发现，不管是接受条件还是拒绝条件下，启动分析思维之后，情感预测偏差都降低了。之所以分析思维与接受/拒绝条件的交互作用显著，这与实验 1 中所采用的情感预测/体验问卷的问法有关。在实验 1 中，1 代表非常不高兴/不开心/不快；3 代表不清楚；5 代表非常高兴/开心/快乐，得分越高表示积极情感越强烈，得分越低则表示消极情感越强烈。在接受条件下，情感预测偏差是正的，数值越大，表示情感预测偏差越大；在拒绝条件下，情感预测偏差是负值，数值越小，表示情感预测偏差越大。在接受条件下，启动分析思维之后，情感预测偏差降低，数值变小了，但是在拒绝条件下，启动分析思维之后，同样是情感预测偏差降低，数值却变大了，因此出现了尽管不管是接受条件还是拒绝条件下，启动分析思维之后，情感预测偏差都降低，二者交互作用显著的结果。

意见 3：实验二被试的招募和测试过程没有介绍得足够清楚。

回应：非常感谢您的建议。我们在实验二中对被试的招募和测试过程进行了补充。

意见 4：实验二的因变量是情感预测，而非影响偏差。从情感预测的差异结果能否推论到影响偏差？可能需要作者给出更多逻辑上的论述以及以往研究的支持。

回应：非常感谢您的建议。的确如您所说实验 2 的因变量是情感预测，而非影响偏差。情感预测影响偏差指的是人们往往高估了未来事件对情感的影响 (Wilson & Gilbert, 2003)。因此，人们对未来事件带来的情感预测强度越高，那么影响偏差可能就会更大；反之，情感预测强度越低，那么影响偏差就会越低。为了检验启动分析思维是否是通过降低了情感预测，进而

降低影响偏差，我们在实验 1 中补充了中介作用分析。中介分析表明，不管是接受条件还是拒绝条件下，情感预测的中介作用都显著。

意见 5: 实验二结果，分析思维对消极情感的预测没有作用，至少从目前已有的数据来看，无法证明作者所认为的“预测生育二孩后的消极情感较少”，因此启动分析思维无效的说法。联系到实验一同样得到分析思维对积极情感与消极情感有不同作用的结果，建议作者可以联系“不同情绪对启发式思维影响”这一主题的研究，可以从理论和假设提出上就考虑到不同情绪效价的影响。

回应: 非常感谢您的建议。实验 2 发现分析思维可以显著降低对生育二孩后积极情感的预测，而没有显著降低生育二孩后消极情感的预测。我们推测，分析思维启动对情感预测的影响可能还会受到任务特征的影响，也可能与情绪效价有关。我们在以后的研究中会针对这一问题继续深入研究。

意见 6: 本研究两个实验都采取了同一种启动分析思维的研究范式，尽管根据以往研究以及预实验的数据可以证明这种方法的操纵有效性，但还是无法排除这一方法引发的其它可能混淆变量的影响，比如思考者图片可能启发了“社会情境”，而几何图片并不会。为提高本研究的内部效度，建议作者补充其它测量或操控分析思维的实验，至少在讨论部分对可能存在的其它解释进行充分讨论。

回应: 非常感谢您的意见。根据您的建议，我们补充了实验 3，采用了语言流畅性任务来启动分析思维，结果发现，采用语言流畅性任务可以有效地启动分析思维，启动分析思维以后显著降低了人们对记忆力测试失败后的消极情感预测和影响偏差，跟实验 1 和实验 2 的结果一致，再次支持了研究假设。

.....

审稿人 3 意见:

意见 1: 本研究考察了分析思维是否会影响情感预测以及情感预测偏差。研究基于双系统理论视角，证实分析 2 思维可以削弱直觉加工，减弱情感预测的强度和预测偏差。本研究对情感预测和情感预测偏差提供了较新的视角，并能够为双系统理论供进一步证据。但是，目前本稿件中还存在一些问题需要作者进一步修改和完善，具体如下。

1. 研究做了非常有趣的实验，但是需要进一步理清研究逻辑。

a) 看到本研究题目后，我的第一反应是“分析思维降低情感预测偏差”，事实上，本研究的假设关注的变量也是“分析思维”和“预测偏差”。从这个角度来说，只有研究 1 证实了本研究的假设。由于研究之间缺乏交互验证，研究证据稍显不足。

b) 作者出假设“分析思维会降低情感预测，进而降低情感预测的影响偏差”。看到这个假设后，我预期本研究会分别考察“分析思维—>情感预测偏差”的直接效应、以及“分析思维—>情感预测—>情感预测偏差”的间接效应。但是，本研究在研究 1 中考察了分析思维分别预测“情感预测”和“情感预测偏差”，研究 2 中考察了“分析思维”预测“情感预测”。从本研究操作定义来看，情感预测是指对于未来事件情感反应的预测 (Wilson & Gilbert, 2003)，情感预测偏差是指预测情感和真实情感之间的差异，这二者并不等同。建议作者进一步补充情感预测和情感预测偏差的相关理论依据以及实验证据。

c) 讨论部分，“总体来看，...启动分析思维可以显著降低人们对未来事件的情感预测，从而降低情感预测的影响偏差”；“本研究发现，分析思维能降低人们对未来事件的情感预测，从而降低情感预测影响偏差”；结论部分“本研究通过两个实验发现分析思维启动会降低人们在决策时的情感预测，并进一步降低情感预测的影响偏差”，这些表述均表明了一种中介关系，

但本研究并未为供实证支持。

回应：a) 非常感谢您的建议。根据您的建议，我们补充了实验 3，采用了不同的分析思维启动方法——语言流畅性任务，进一步考察了分析思维启动对情感预测和影响偏差的作用，结果发现，分析思维启动可以显著降低情感预测和影响偏差，再次证实了实验 1 和实验 2 的研究结果，提供了进一步的证据。

b) 情感预测影响偏差指的是人们往往高估了未来事件对情感的影响(Wilson & Gilbert, 2003)。因此，人们对未来事件带来的情感预测强度越高，那么影响偏差可能就会更大；反之，情感预测强度越低，那么影响偏差就会越低。我们在研究假设提出部分补充了情感预测与情感预测偏差的关系，详见正文 1 引言部分。为了检验启动分析思维是否是通过降低了情感预测，进而降低影响偏差，我们在实验 1 中补充了中介作用分析。中介分析表明，不管是接受条件还是拒绝条件下，情感预测的中介作用都显著。

c) 我们已经在实验 1 中补充了中介效应检验，证明了情感预测在启动分析思维降低影响偏差的过程中起中介作用。

意见 2：请清晰的描述研究步骤和细节。

a) 两个实验对分析思维进行操纵，是否有操纵检查？

b) 预实验、实验 1 中各分组有多少人？

c) 实验 1 中，被试在启动分析思维后，填写“提议被对方接受或者拒绝后的情感”，那么被试是只评定一种情况还是两种情况都评定？看表 2 呈现的数据，仅报告了被试评定一种情况的情绪预测。

d) 实验 2 和实验 3 中的图片呈现方式是不是一致的？如果是一致的，请注明同预实验。

回应：a) 实验 1 和实验 2 中采用图片启动分析思维，实验 3 采用语言流畅性任务启动分析思维。为了检验启动是否有效，我们分别进行了预实验。实验 1 中预实验检验了图片启动分析思维的操纵有效性，实验 3 中预实验检验了语言流畅性任务启动分析思维的有效性。结果表明，两种方法都是有效的。因此，在实验 1 正式实验、实验 2、实验 3 正式实验中都没有再进行操纵检验。这样设计是出于以下考虑：如果在正式实验中再次进行操纵检验，启动组和控制组都需要完成 CRT，而 CRT 任务有可能会干扰启动效果。

b) 我们已经在正文补充相应内容。

c) 情感预测时是两种情况都预测，情感体验时是根据结果只评定一种情况。我们已经在实验 1 流程中进行了补充。

d) 实验 2 和实验 3 中的图片呈现方式是一致的，已经在正文中注明同预实验。

意见 3：作者在自检报告中回答“本研究采用实验法，不存在共同方法偏差”。事实上，当被试因为同样的测量环境、施测语境以及研究设计本身特征所造成的测量反应偏差时，就会发生共同方法偏差。换句话说，心理测量通过对测验项目所进行的行为反应间接推断个体的心理属性，不可避免会存在共同方法偏差问题。因此，您研究的结果同样会受到共同方法偏差的影响。建议进行共同方法偏差检验，最简单方法是采用 Harman 单因子法。

回应：感谢您的建议。根据周浩和龙立荣（2004），共同方法偏差指的是因为同样的数据来源或评分者、同样的测量环境、项目语境以及项目本身特征所造成的预测变量与效标变量之间人为的共变。根据朱海腾和李川云（2019），如果两个变量用同一方法测量或测量方法有某些共同之处(例如数据来自同一受测者)，就会共享一部分方法变异，形成共同方法变异。共同方法偏差是由此衍生的概念，指观测相关系数偏离真实相关系数的程度，多数情况下表现为观测相关系数的膨胀或高估，有时可能引起假阳性结果，导致错误的因果关系推论。在

本研究的3个实验中，预测变量全都是实验操纵的，效标变量（即情感预测、情感体验）则是测量的。预测变量和效标变量没有共同的数据来源或评分者等，根据朱海腾和李川云（2019），本研究共同方法偏差的风险极低。

参考文献：

周浩，龙立荣. (2004). 共同方法偏差的统计检验与控制方法. 心理科学进展, 12(6), 942-950.
朱海腾，李川云. (2019). 共同方法变异是致命瘟疫吗？-论争、新知与应对. 心理科学进展, 27(4), 587-599

意见 4：本研究中计划样本量分别采用了 $d = 0.8$, $f = 0.25$ 。选择这两个效应量的依据是什么？是基于以往研究的结果还是其他依据？建议说明。

回应：感谢您的建议，我们根据以往研究估计的效应量。我们在正文中已经补充。

意见 5：本研究中全部采用自编问卷，请报告每个量表的题目数、量表信度。此外，如果问卷较短，建议将问卷条目直接添加在“方法”部分；如果问卷条目较多，建议将问卷条目作为附录添加在文章 7 中。

回应：感谢您的建议，由于本研究中情感预测和体验问卷都只有 3 个词语，我们已经将问卷条目补充在正文中。同时也补充了问卷信度系数。

意见 6：实验 1 中采用了积极情绪词作为测评工具，并没有消极情感词。因此，建议在结果报告时不要使用“消极情感”等类似表述，采用表 3 中的“情感预测”或者“情感体验”

回应：感谢您的建议。在情感预测的研究中，不管对积极事件还是消极事件的预测，研究者通常采用 how happy/pleasant 来提问，让被试进行情感预测。例如，在 Gilbert 等人（1998）关于情感预测的经典研究中，让那些没有经历过分手的被试预测分手后的情感“how happy they would be 2 months after experiencing such a dissolution”，评分采用 7 点量表，从 1（not happy）到 7（very happy）。Hoerger 等人（2016）的研究中，让被试对支持的球队赢球或输球后的情感进行预测，“how happy they would feel the following Monday both in the event of a win and a loss”。借鉴前人的做法，在实验 1 中，我们要求被试对接受或拒绝后的情感进行预测都采用了同样的问题。尽管问卷中只采用了正效价的情绪词（高兴、开心、快乐），然而被试在评价的时候却包含了消极情绪，采用 5 点量表，1 代表非常不高兴/不开心/不快乐；5 代表非常高兴/开心/快乐，得分越高表示积极情感越强烈，得分越低则表示消极情感越强烈。

第二轮

审稿人 1 意见：

意见 1：作者对“实验 1 选取的三种情绪词(高兴、开心、快乐)”这个问题的回答，可能涉及到文化差异。how happy 并不是有多高兴。更好的翻译方法，应该问被试：“您(不)高兴的程度如何？”从 1 非常不高兴；到 4 没有不高兴，也没有高兴；到 7 非常高兴。可能会是更好的测量。发表出来的文章，也并不是十全十美，也仍有改进的空间。建议作者以后的研究中，可以在 pilot study 中在 debriefing 的环节，直接问问被试作答的感受。毕竟问：how happy are u today? 在英文环境中是个很平常的问题，但问你有多高兴，这个实际上不是中文环境中的一个平常问题。当然，这是一个非常非常小的问题啦。

回应：非常感谢您的建议，我们在以后的研究中会注意这个问题。

审稿人3意见:

意见1: 经过一轮的修改, 作者就审稿专家的意见给予了充分的回应; 使得文章结构更加清晰; 增加了一个实验也一定程度上提高了研究的内部效度。但同时也使一些令人疑惑的问题凸显出来, 请作者进一步澄清:

1. 还是双系统理论的问题。作者列举了以往干预情感预测影响偏差 (affective forecasting impact bias, 以下简称 AFIB) 的三种方法: 去焦点化训练、心理模拟理论和正念训练。为什么之前这些干预方法都没有基于双系统理论而提出; 或者反过来说, 作者引入双系统理论在解释 AFIB 的作用机制上, 可以提供哪些独特贡献?

回应: 非常感谢您的意见, 这促使我们深入思考研究的价值。以往研究对情感预测影响偏差的干预进行了多种尝试, 不同的尝试方法基于不同的影响偏差来源的假设。例如, 去焦点化训练 (defocusing exercise), 基于影响偏差来源于焦点错觉 (focus illusion) 的假设, 即影响偏差是因为在情感预测时只关注中心事件的影响力, 而忽视了与之伴随的偶然事件的影响; 适应训练, 基于影响偏差来源于免疫忽视 (immune neglect) 的假设, 即在情感预测时未能考虑人们会适应将来的事件; 代理人策略 (surrogation), 基于心理模拟理论, 即影响偏差是因为人们在预测时对未来模拟不准确; 正念训练, 也是基于焦点错觉的假设, 通过正念训练提升人们对情绪的理解和觉察, 更注意细节, 从而降低情感预测的强度, 降低影响偏差。本研究提出的分析思维启动方法则是基于影响偏差来源于预测时所基于的直觉信念的假设。通过启动分析思维来干预预测时人们使用的直觉信念, 进而降低情感预测强度, 降低影响偏差。以往几乎没有见到有研究从双系统理论出发, 通过启动分析思维来干预直觉信念, 进而减少影响偏差。

本研究从双系统理论视角出发, 考察分析思维对情感预测偏差的干预作用, 结果发现分析思维能降低人们对未来事件的情感预测, 从而降低情感预测影响偏差, 这为情感预测影响偏差的干预提供了一个新的研究方向。另外, 本研究表明启动分析思维后, 系统2可以压倒或削弱系统1的直觉加工, 为双系统理论提供了进一步证据。

意见2: 还是情感预测与情感预测偏差两变量之间关系的问题。原稿中实验二用情感预测来代替影响偏差, 显然是不正确的做法。但现稿将验证模型改为“分析思维→情感预测→情感预测偏差”, 又会导致如下其他问题。(1) 整个研究重点与框架需要调整。需要重点阐述情感预测的解释意义。情感预测偏差=情感预测-真实情感体验, 作者是想证明: 分析思维是通过改变情感预测, 而非真实情感, 从而影响预测偏差? 那么需要对目前的文献综述重新进行调整与补充, 增加分析思维对情感预测影响的理论与研究的介绍。(2) 实验一的中介只是部分中介, 那么是否分析思维也影响到了真实情感, 实验一需要补充关于真实情感的结果。

(3) 实验三同样需要补充“分析思维→情感预测→情感预测偏差”中介分析, 以及关于真实情感的结果。(4) 关于研究意义, 如果发现分析思维其实改变的是情感预测, 那么还有没有必要提预测偏差呢? 毕竟情感预测偏差=情感预测-真实情感体验, 偏差结果相应发生改变是很显然的结论。

回应: (1) 感谢您的建议, 您建议我们以情感预测为重点重新组织文献综述。所谓情感预测影响偏差指的是情感预测时人们经常会高估未来事件对其情感的影响。因此, 情感预测影响偏差包含情感预测, 是其题中应有之义。当前学界关于情感预测偏差的研究与情感预测的研究并没有明显区分, 情感预测影响偏差的研究综述已经包含了情感预测。鉴于此, 我们保留了目前的研究框架, 仍然以情感预测偏差为重点。根据您的建议, 我们在文献综述部分补充了相应的内容, 进一步强调了情感预测在影响偏差干预中的作用。

(2) 感谢您的建议。我们在实验一中补充了情感体验的中介分析。

(3) 我们在实验三中补充了情感预测和情感体验的中介作用分析, 结果发现, 情感预测在分析思维启动降低影响偏差中起到了中介作用, 而情感体验的中介作用不显著。

(4) 本研究关注的问题是如何降低情感预测影响偏差, 分析思维会通过改变情感预测的强度进而改变影响偏差。以往关于情感预测偏差的干预研究大多数也是通过改变情感预测的强度来对影响偏差进行干预的。因此, 我们认为还是有必要提影响偏差的。

意见 3: 还是实验一结果中分析思维与两种方案的交互作用。我可以理解方向上的差异, 但偏差程度上(偏差绝对值)积极与消极情感是否存在差异?

回应: 感谢您的建议。我们对偏差取绝对值后进行了分析, 结果发现, 拒绝条件下(消极情感)预测的偏差程度大于接受条件(积极情感)。我们在正文中补充了相应的内容。

意见 4: 从被试性别比例上看, 实验一男女比 92 vs. 148, 实验二全部女性被试, 实验三男女比 9 vs.43。性别比例严重失衡, 需要补充是否存在性别差异及其影响的结果。

回应: 非常感谢您的建议。我们在文中补充了性别差异分析。实验一中男女被试的情感预测影响偏差不存在显著差异。在正文 2.2.1 被试 部分进行了相应补充, 具体内容如下: 男生(M 男 = 0.79, SD 男 = 0.91)和女生(M 女 = 0.72, SD 女 = 0.61)的影响偏差不存在显著差异, $t(238) = 0.71, p = 0.48, 95\%$ 置信区间[-0.12, 0.40]。

实验二因为是对生育二孩的情感进行预测, 因此全部选择了女性被试。

实验三中, 男女被试的情感预测影响偏差也不存在显著差异。在正文 4.2.1 被试 部分进行了补充, 具体内容如下: 在本实验中, 男生(M 男 = 0.74, SD 男 = 1.21)和女生(M 女 = 0.42, SD 女 = 0.65)总的影响偏差不存在显著差异, $t(50) = 1.14, p = 0.26, 95\%$ 置信区间 [-0.24, 0.89]。

第三轮

审稿人 3 意见: 作者已对所有提出问题给出了认真可靠的回应。同意发表。

编委复审意见: 综合考虑审稿人的意见, 编委同意发表。

主编终审意见:

目前正文约 14000 字, 前言介绍部分也确实显得啰嗦、堆砌文献, 没有做到“经分析现状、存在问题后, 找出本文要解决或回答的科学问题”, 亦正是如此, 读者很难看出其研究意义。建议作者做精简(不是单纯字数的删减)修改。

回应: 非常感谢您的建议。我们已经按照“经分析现状、存在问题后, 找出本文要解决或回答的科学问题”的逻辑精简了文献综述, 并进一步精简了全文, 目前正文字数为 10700 字。