

## 《心理学报》审稿意见与作者回应

题目：基于不同冲突水平的认知控制对性别刻板印象表达的影响

作者：陈莉 石晓柯 李维娜 胡妍

---

### 第一轮

#### 审稿人 1 意见：

《基于不同冲突水平的认知控制对性别刻板印象的影响》该研究基于经典的认知控制加工理论和冲突监控理论，探究不同冲突水平的任务背景下，不同条件诱发的认知控制对性别刻板印象表达的影响。结果发现：(1) 当目标信息为图像刺激时，不同冲突比例的任务水平对性别刻板印象的表达没有差异；(2) 而当目标信息为语义刺激时，高冲突比例的任务水平易抑制性别刻板印象的表达，低冲突比例的任务水平易加强性别刻板印象的表达，并且高冲突水平下产生的认知控制加工能够在随后不同的任务中得到保持。作者得出结论不同冲突比例水平的任务引发的认知控制可以抑制性别刻板印象的表达，但会受到信息呈现形式的影响。本文进行三个系列实验，研究问题较为新颖，有一定研究意义。然而也有一些不足待修正：1 前言部分研究者提到“虽然不同的反应控制均提供了一种调节偏差的机制，但因为这种控制模式只在偏差出现或激活之后才开始发挥作用，所以人们在完成任务之前往往不能及时纠正其存在的偏见反应(Amodio & Swencionis, 2018)。就本质特点而言，利用反应性控制的方法来削弱刻板印象并不会起到持久的干预效果。”但从当前的研究中，作者并没有比较反应性控制和主动性抑制在治疗上的持久性差异问题。2 前言部分假设 1 和假设 2 有什么区别？假设写得过于具体，更像是对其结果的具体预期。假设应该是本研究想解决核心问题的几条可能结果的概括性预测。此外，为什么高低冲突下，参与者的成绩与基线相比是相反的模式？作者在前面没有铺垫出来，因此这里的为什么有这种预期很难推理出来。3 方法部分有些描述不清楚。比如“单样本 t 检验的结果表明，男性图片和女性图片与较极端的面孔整体印象水平均有显著差异( $p < 0.001$ )。”这个比较的目的是什么？极端的面孔整体印象水平是什么意思？为什么要进行这个比较？“单样本 t 检验的结果表明，工作图片和家庭图片的代表性均与图片代表性的中间水平差异显著( $p < 0.001$ )。”这里也同样，比较的目的是什么？4 方法部分，高冲突、基线和低冲突以不一致诗词呈现比例的从高到低界定是否有依据？还是仅仅可以称为冲突多、冲突中、冲突少？5 结果部分，尽管实验 1 没结果，也应该把每种条件下的反应时和正确率列出来，否则后面的讨论读者没有看到数据，无法评

估作者陈述的可信性。6 结果部分，实验 1 的阴性结果是否有另一种可能性，在该范式中，目标刺激是图片，但是不同于以往范式，目标出现的时候没有让被试按键，而是掩蔽刺激出现后才让被试按键。由于人们对图片的加工很快，是不是在三类冲突的条件下，目标呈现的时候就对冲突刺激做好了加工，掩蔽刺激出现时仅仅是做出反应，因此掩盖了不同冲突类型带来的反应的差异。7 结果部分，实验 2 和实验 3 低冲突和基线没有统计上的差别，就得出低冲突水平上表现最差这一结论。没有主效应和交互作用，却得出“不过数据显示，PDP-A 的估算值在高冲突条件下的自动估算值最高 ( $M=0.69, SD=0.05$ )，基线水平位于中间 ( $M=0.63, SD=0.05$ )，低冲突水平最低 ( $M=0.56, SD=0.06$ )”这一结论，是不可信的。此外，这两个实验都存在速度和正确率的权衡问题，所以应该权衡反应时和正确率两个指标进行统计，因此，目前的结论尚不可靠。

**意见 1：**前言部分研究者提到“虽然不同的反应控制均提供了一种调节偏见的机制，但因为这种控制模式只在偏见出现或激活之后才开始发挥作用，所以人们在完成任务之前往往不能及时纠正其存在的偏见反应(Amodio & Swencionis, 2018)。就本质特点而言，利用反应性控制的方法来削弱刻板印象并不会起到持久的干预效果。”但从当前的研究中，作者并没有比较反应性控制和主动性抑制在治疗上的持久性差异问题。

**回应：**非常感谢审稿专家对文章的细致分析。根据审稿专家的宝贵意见，再次认真阅读前人研究，并深入思考本文的研究意义和主要贡献，本次修改对原文引言中的部分表述进行了重新书写。为方便审稿专家审阅，本次修改补充的内容均用红色字体对原文相应内容进行标识。具体修改内容如下：

认知控制指的是人们能根据目标或任务做出相应的反应，遇到分心物能保存目标、抑制习惯性反应或冲动行为的能力(Miller & Cohen, 2001)，是目标导向行为的基础，在实验室条件下会采用冲突任务来研究认知控制(Gonthier & Ambrosi, 2021)。冲突监控理论(conflict monitoring theory; Veen & Carter, 2005)是最常见的对冲突任务的解释。该理论认为，冲突受到自上而下认知控制的调节，当先前的试次是不一致试次会引发调整机制，认知控制水平提高，但当先前试次为一致试次，则无需调整，即个体在先前试次的冲突检测导致在当前试次更好地控制了冲突(Botvinick et al., 2001; Botvinick et al., 1999)。大量研究表明认知控制分为两类，对快速变化的外部条件作出即时行动一般属于认知控制过程中的反应性控制，对快速变化的外部条件做出更为持久的战略优化行动一般属于认知控制过程中的主动控制(Funes et al., 2010; Braver et al., 2009)。例如，一个人在下班后总是在十字路口左转，但听说

那条路上堵车，他可能在到达十字路口之前提醒自己右转，这就是主动控制；但他也可能采用反应控制模式，即到达十字路口后，遇见堵车，然后右转以避开拥挤的交通。此时，反应控制比主动控制需要的认知资源少，但更容易出错。

迄今为止，几乎所有关于控制刻板印象或偏见的研究都集中在反应控制上。反应控制是在刻板印象的被检测和发现后，对其进行反应性的抑制(Mark & Muraven, 2008)。反刻板样例暴露法(Ramasubramanian, 2015)、生动反刻板印象情景故事法(Burn et al., 2017)、虚假 IAT 结果反馈(Cvencek et al., 2010)等方法均属于反应控制的干预训练法。以往的研究表明，这些方法能够暂时有效地削弱刻板印象带来的影响(Lai et al., 2016)。然而，尽管反应性控制为克服偏见提供了一个重要的机制，它只在偏见出现或激活之后才开始发挥作用，因此容易受到社会中各种情境因素的不断强化以及认知资源的影响，从而破坏它的有效性，出现干预效果消退的情况(Lambert et al., 2003; Calvin, 2016)。

不同于反应性控制，主动控制是一种完全不同的干预概念。作为一种目标相关的控制方式，它针对的是一个人的预期反应，而不是偏见的来源。主动控制采用了操纵任务干扰的方法，它可以在激活人们头脑中关于刻板印象的相关信息之前，有效地将刻板印象的关注点迁移至任务目标本身，通过对任务困难的体验或对一个人的预期行为增加的动机来限制内隐偏见在行为中表达的可能性(Amodio et al., 2018)。已有研究表明，在种族分类任务中观察到的刻板印象偏见的反应主要是由冲突驱动的，即高冲突增强了被试对任务目标的关注，可能导致积极主动的控制模式(Bartholow & Dickter, 2008; Dickter & Bartholow, 2010)，但这一研究并未直接验证主动控制对种族偏见的干预。在一项最新研究中(Amodio et al., 2018)，研究者提出了主动控制模型(Proactive Model of Control)(图 1)，采用了武器识别范式(Payne, 2001)、偏见启动任务与刻板印象启动任务考察了主动控制对内隐偏见和内隐刻板印象表达的抑制作用，首次证实了相比基线组，个体在高冲突水平下会采用主动控制的加工模式，抑制偏见或者刻板印象的表达，在低冲突水平下则会更明显地表现出内隐种族偏见及内隐种族刻板印象。然而，以上研究中，主要是针对种族相关的偏见行为。然而，在现实生活中，与性别相关的刻板印象与每一个个体都息息相关，尤其体现在男性和女性更适合从事哪一类职业的固定观念上，而这种观念会导致职业性别产生隔离(Gross, 1968)，从而对不同行业的就业产生消极影响(Koenig & Eagly, 2014)。遗憾的是，主动控制对性别刻板印象的影响并未引起研究者的广泛关注，所以，这一认知加工方式是否能同样有效地抑制性别刻板印象还有待探究。

新增参考文献：

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control

within human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), 7351–7356.

Gonthier, C., Ambrosi, S., & Blaye, A. (2021). Learning-based before intentional cognitive control: developmental evidence for a dissociation between implicit and explicit control. *Journal of Experimental Psychology Learning Memory and Cognition*.

Gross, E.(1968). Plus Ca Change...? The Sexual Structure of Occupations Over Time. *Social Problems* 16, 198-208.

Koenig, A. M., Eagly, A. H. 2014. “Extending role congruity theory of prejudice to men and women with sex-typed mental illnesses.” *Basic & Applied Social Psychology*, 36(1), 70-82.

Veen, V. V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI Study. *NeuroImage*, 27(3), 497–504.

**意见 2:** 前言部分假设 1 和假设 2 有什么区别？假设写得过于具体，更像是对其结果的具体预期。假设应该是本研究想解决核心问题的几条可能结果的概括性预测。

**回应:** 非常感谢审稿专家对文章的细致分析。根据审稿专家的宝贵意见，对研究假设进行了修改。

具体修改内容如下：

目前，较少有研究直接涉及到不同冲突水平的认知控制对性别刻板印象的影响。针对上述问题，本研究假设：(1)在图片分类-确认实验任务中，与基线水平相比，被试在高冲突水平下驱动更强的主动控制调整，性别刻板印象的表达更容易被抑制。(2)在词语分类-确认实验任务中，与基线水平相比，被试在高冲突水平下驱动更强的主动控制调整，性别刻板印象的表达更容易被抑制。(3)在先前不同冲突水平的词语分类任务后，高冲突水平下的反应更能抑制后续刻板 Flanker 任务性别刻板印象的表达，驱动更强的主动控制。

**意见 3:** 方法部分有些描述不清楚。比如“单样本 t 检验的结果表明，男性图片和女性图片与较极端的面孔整体印象水平均有显著差异( $ps < 0.001$ )。”这个比较的目的是什么？极端的面孔整体印象水平是什么意思？为什么要进行这个比较？“单样本 t 检验的结果表明，工作图片和家庭图片的代表性均与图片代表性的中间水平差异显著( $ps < 0.001$ )。”这里也同理，比较的目的是什么？

**回应:** 非常感谢审稿专家对文章的细致分析。根据审稿专家的宝贵意见，本次修改补充的内

容均用蓝色字体对原文相应内容进行标识。

具体修改内容如下：

本研究中男女图片采用 7 点评分，“1”代表极具女性化，“7”代表极具男性化，而研究中希望男性图片和女性图片是中性水平，即中间值“4”，所以是用男性图片、女性图片的均值分别与极端值“7”和“1”比较，如果差异显著则表明图片并不极端，符合中性水平。

而家庭图片和工作图片是采用 5 点评分，“1”代表非常不符合家庭或者工作图片，“5”代表非常符合家庭或工作图片，而研究中希望家庭和工作的图片选择要接近“5”这一极端值，所以用家庭图片、工作图片的均值与中性水平“3”比较，如果差异显著，则说明选择的图片均显著大于 3。

参考文献：

崔诣晨. (2016). *知觉冲突视阈下印象形成的认知控制* (博士学位论文). 上海师范大学.

**意见 4：**方法部分，高冲突、基线和低冲突以不一致试词呈现比例的从高到低界定是否有依据？还是仅仅可以称为冲突多、冲突中、冲突少？

**回应：**非常感谢审稿专家对文章的细致分析。根据审稿专家的宝贵意见，本次修改补充的内容均原文相应内容进行标识。

具体修改内容如下：

从本质上来说，认知控制任务中，当任务比较容易时，参与者更依赖反应控制模式，只有遇到干扰物才会进行控制；而当任务困难时，参与者更加依赖积极主动地控制模式，在遇到干扰物前，将注意力集中在任务目标中。因此，增加不一致试次的频率会增加对主动控制的需求。而冲突适应理论为主动控制提供了早期依据(Gratton et al., 1992)。例如，Gratton 等人证明，当任务干扰程度较低时（例如，只有 25% 不一致试验中），观察到强烈的侧翼不相容效应，这表明干扰物的强烈作用和反应性控制模式；而当干扰程度较高时（例如，75% 的试验不一致），侧翼不相容效应显著降低，表明干扰物的影响降低，控制模式更主动。因此，本研究沿用 Amodio（2018）的研究采用不同的冲突试次频率，高冲突（80%）、基线水平（50%）以及低冲突（20%），验证高冲突条件下，参与者采用积极主动地控制模式，增强对任务的关注，从而抑制性别刻板印象的表达；而低冲突条件下，减少主动控制，依然存在明显的性别刻板印象。

新增参考文献：

Gonthier, C., Ambrosi, S., & Blaye, A. (2021). Learning-based before intentional cognitive control: developmental

evidence for a dissociation between implicit and explicit control. *Journal of Experimental Psychology Learning Memory and Cognition*.

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121(4), 480–506.

**意见 5:** 结果部分，尽管实验 1 没结果，也应该把每种条件下的反应时和正确率列出来，否则后面的讨论读者没有看到数据，无法评估者陈述的可信性。

**回应:** 非常感谢审稿专家提出的问题，我们对结果数据进行了补充。具体修改内容如下：

冲突水平	启动刺激 × 目标刺激	正确率 <i>M(SE)</i>	反应时 <i>M(SE)</i>	总正确率 <i>M(SE)</i>	总反应时 <i>M(SE)</i>
高	男-工作	0.96(0.01)	452.89 (14.63)	0.96 (0.01)	451.74 (12.90)
	男-家庭	0.94(0.01)	451.04 (13.24)		
	女-工作	0.96(0.01)	451.97 (13.54)		
	女-家庭	0.96(0.01)	451.05 (14.07)		
基线	男-工作	0.96(0.01)	456.01 (13.58)	0.95 (0.01)	455.46 (12.38)
	男-家庭	0.94(0.01)	461.46 (13.44)		
	女-工作	0.95(0.01)	456.98 (13.31)		
	女-家庭	0.94(0.01)	447.39 (12.37)		
低	男-工作	0.96(0.01)	453.36 (14.49)	0.95 (0.01)	459.34 (13.97)
	男-家庭	0.95(0.01)	459.01 (14.86)		
	女-工作	0.95(0.01)	466.83 (15.21)		
	女-家庭	0.95(0.01)	458.16 (15.20)		

	主动控制 <i>M(SE)</i>	反应控制 <i>M(SE)</i>
高冲突	0.92 (0.01)	0.79 (0.06)
基线	0.90 (0.01)	0.49 (0.06)
低冲突	0.91 (0.01)	0.42 (0.06)

**意见 6:** 结果部分, 实验 1 的阴性结果是否有另一种可能性, 在该范式中, 目标刺激是图片, 但是不同于以往范式, 目标出现的时候没有让被试按键, 而是掩蔽刺激出现后才让被试按键。由于人们对图片的加工很快, 是不是在三类冲突的条件下, 目标呈现的时候就对冲突刺激做好了加工, 掩蔽刺激出现时仅仅是做出反应, 因此掩盖了不同冲突类型带来的反应的差异。

**回应:** 感谢审稿专家提出的问题。我们对您的疑惑解释如下:

实验 1 的流程是在空屏后, 会呈现一张男或女的面孔图片 (200ms), 紧接着会呈现一张场景图片 (200ms), 最后会呈现一张掩蔽刺激的图片, 被试需要在掩蔽刺激呈现时对场景图片进行分类。首先, 实验 1 和实验 2 是在相同流程中进行, 均是在男女性别面孔呈现 200ms 后出现目标刺激, 这样做是为了客观地比较不同目标刺激的呈现方式, 即图片和文字刺激, 在不同冲突水平下, 是否会同样抑制性别刻板印象的表达, 驱动更强的主动控制调整; 其次, 以往研究中也采用了相似的任务流程, 例如, 在 Chen & Zeng (2021) 的研究中, 采用图片-分类确认任务, 启动刺激为男性或者女性图片, 目标刺激为笔或者注射器的图片, 然后呈现掩蔽刺激, 被试需要在掩蔽刺激呈现时对笔和注射器作出反应, 这一实验验证在消极情绪状态下, 主动控制对内隐刻板印象的影响; 最后, 瞬时记忆时间短, 而且容量大, 很容易形成对图像信息的记忆, 所以存在目标呈现时被试已经做好了加工, 等待掩蔽刺激出现后进行反应。因此, 后续研究也可以缩短掩蔽刺激呈现的时间或者直接让被试在目标刺激出现时作反应, 探究是否会抑制刻板印象的表达以及主动控制方式的出现。

**参考文献:**

Chen, L., Zeng, S, Y., & He, R. (2021). Negative emotions can interfere with the inhibitory effect of proactive control on implicit stereotypes. *North American Academic Research Journal*, 4(4),53-69.

**意见 7:** 结果部分, 实验 2 和实验 3 低冲突和基线没有统计上的差别, 就得出低冲突水平上表现最差这一结论。没有主效应和交互作用, 却得出“不过数据显示, PDP-A 的估算值在高冲突条件下的自动估算值最高(M=0.69, SD=0.05), 基线水平位于中间(M=0.63, SD=0.05), 低冲突水平最低(M=0.56, SD=0.06)”这一结论, 是不可信的。此外, 这两个实验都存在速度和正确率的权衡问题, 所以应该权衡反应时和正确率两个指标进行统计, 因此, 目前的结论尚不可靠。

**回应:** 感谢审稿专家提出的具体意见。针对审稿专家提出的问题, 通过查阅相关文献, 我们再次梳理了原文的数据统计方法及其结果表述, 对此问题做出如下回应:

实验 2 和实验 3 低冲突水平和基线水平没有显著差异, 同时, PDP-A 的估算值在冲突

水平和启动刺激未出现主效应以及交互作用,因此均不能进行比较,原文中已删除此处描述。  
感谢审稿专家的细心审阅。

针对速度-准确性权衡的问题,在高冲突条件下,由于不一致试次的频率增多,产生冲突适应(对先前试次的冲突监控导致在当前试次更好地控制冲突),增加对主动控制的需求,参与者注意力更好的聚焦在任务本身,所以反应时增加,准确率提高。而低冲突条件下,不一致试次只有 20%,任务比较容易,所以被试的反应时间变短,注意力不能高度集中,这样准确率很低。

.....

#### 审稿人 2 意见:

《基于不同冲突水平的认知控制对性别刻板印象的影响》选题具有重要的理论与现实意义,为性别刻板印象的干预提供了主动控制的策略。论文总体研究很扎实,目前行文上主要有如下方面值得进一步完善: 1、引言中对于研究的理论基础认知控制加工理论和冲突监控理论有待更多提及,同时,三个研究假设的提出逻辑有待进一步做好铺垫,需要有理有据; 2、对于研究中所采用的图片材料和文字语义材料,这些刺激材料可能产生不同结果以及理论需要在问题提出中补充说明; 3、研究 2 和研究 3 之间的逻辑需要更清晰地说明,从实验材料和范式来看,这两者的递进关系不明显; 4、对于文中一些术语的表达,如冲突条件,刻板不一致条件等有待区分,此外,“内隐”性别刻板印象的内隐如何体现也有待交代; 5、总讨论部分略显笼统,对于本研究的主要理论创新和贡献有待丰富和拓展,建议采用小标题更有层次地论述。 6、参考文献的引用有待进一步符合《心理学报》的最新规范。

**意见 1:** 引言中对于研究的理论基础认知控制加工理论和冲突监控理论有待更多提及,同时,三个研究假设的提出逻辑有待进一步做好铺垫,需要有理有据;

**回应:** 感谢审稿专家提出的意见。根据审稿专家的宝贵意见,对引言部分进行了补充修改。具体修改如下,用红色字体标出:

认知控制指的是人们能根据目标或任务做出相应的反应,遇到分心物能保存目标、抑制习惯性反应或冲动行为的能力(Miller & Cohen, 2001),是目标导向行为的基础,在实验室条件下会采用冲突任务来研究认知控制(Gonthier & Ambrosi, 2021)。冲突监控理论(conflict monitoring theory; Veen & Carter, 2005)是最常见的对冲突任务的解释。该理论认为,冲突受到自上而下认知控制的调节,当先前的试次是不一致试次会引发调整机制,认知控制水平提



高，但当先前试次为一致试次，则无需调整，即个体在先前试次的冲突检测导致在当前试次更好地控制了冲突(Botvinick et al., 2001; Botvinick et al., 1999)。大量研究表明认知控制分为两类，对快速变化的外部条件作出即时行动一般属于认知控制过程中的反应性控制，对快速变化的外部条件做出更为持久的战略优化行动一般属于认知控制过程中的主动控制(Funes et al., & Humphreys, 2010; Braver et al., 2009)。例如，一个人在下班后总是在十字路口左转，但听说那条路上堵车，他可能在到达十字路口之前提醒自己右转，这就是主动控制；但他也可能采用反应控制模式，即到达十字路口后，遇见堵车，然后右转以避开拥挤的交通。此时，反应控制比主动控制需要的认知资源少，但更容易出错。

迄今为止，几乎所有关于控制刻板印象或偏见的研究都集中在反应控制上。反应控制是在刻板印象的被检测和发现后，对其进行反应性的抑制(Mark & Muraven, 2008)。反刻板样例暴露法(Ramasubramanian, 2015)、生动反刻板印象情景故事法(Burn et al., 2017)、虚假 IAT 结果反馈(Cvencek et al., 2010)等方法均属于反应控制的干预训练法。以往的研究表明，这些方法能够暂时有效地削弱刻板印象带来的影响(Lai et al., 2016)。然而，尽管反应性控制为克服偏见提供了一个重要的机制，它只在偏见出现或激活之后才开始发挥作用，因此容易受到社会中各种情境因素的不断强化以及认知资源的影响，从而破坏它的有效性，出现干预效果消退的情况(Lambert et al., 2003; Calvin, 2016)。

不同于反应性控制，主动控制是一种完全不同的干预概念。作为一种目标相关的控制方式，它针对的是一个人的预期反应，而不是偏见的来源。主动控制采用了操纵任务干扰的方法，它可以在激活人们头脑中关于刻板印象的相关信息之前，有效地将刻板印象的关注点迁移至任务目标本身，通过对任务困难的体验或对一个人的预期行为增加的动机来限制内隐偏见在行为中表达的可能性(Amodio et al., 2018)。已有研究表明，在种族分类任务中观察到的刻板印象偏见的反应主要是由冲突驱动的，即高冲突增强了被试对任务目标的关注，可能导致积极主动的控制模式(Bartholow & Dickter, 2008; Dickter & Bartholow, 2010)，但这一研究并未直接验证主动控制对种族偏见的干预。在一项最新研究中(Amodio et al., 2018)，研究者提出了主动控制模型(Proactive Model of Control)(图 1)，采用了武器识别范式(Payne, 2001)、偏见启动任务与刻板印象启动任务考察了主动控制对内隐偏见和内隐刻板印象表达的抑制作用，首次证实了相比基线组，个体在高冲突水平下会采用主动控制的加工模式，抑制偏见或者刻板印象的表达，在低冲突水平下则会更明显地表现出内隐种族偏见及内隐种族刻板印象。然而，以上研究中，主要是针对种族相关的偏见行为。然而，在现实生活中，与性别相关的刻板印象与每一个个体都息息相关，尤其体现在男性和女性更适合从事哪一类职业的固

定观念上，而这种观念会导致职业性别产生隔离(Gross, 1968)，从而对不同行业的就业产生消极影响(Koenig & Eagly, 2014)。遗憾的是，主动控制对性别刻板印象的影响并未引起研究者的广泛关注，所以，这一认知加工方式是否能同样有效地抑制性别刻板印象还有待探究。

从刻板印象的图式化语义表征特点出发，以具体图像信息作为考察刻板印象认知加工的目标刺激，其研究结果对刻板印象加工机制的解释不可避免存在一定的局限(王沛等, 2010)。更为重要的是，就信息的承载形式而言，知觉者在社会生活中常借助以词语或句子为形式的语义信息来进行日常交流，若是以被赋予一定社会属性的图片信息作为目标刺激，不论所选取的信息与某一群体的联系有多紧密(例如，男性-工作场景，女性-家庭场景等)，都无法完全表现出对某一群体刻板印象概念化的抽象表征。例如，Chen & Zeng (2021) 的研究中，采用图片分类确认方式，得出在消极情绪下主动控制模式更多的出现在高冲突水平中，但是语义加工是一种自上而下的加工，而且更加抽象，因此不同信息的呈现方式是否对刻板印象的激活和表达，以及这两种表征方式是否有助于启动主动控制策略，有所不同需要进一步研究。

随着刻板印象研究的深入，关于不同冲突比例水平下的认知控制的基础研究已经取得了相当大的进展(李政汉等, 2018)，但在不同的冲突比例水平下，认知控制对图片刺激所诱发的性别刻板印象表达会有着怎样的影响呢？此外，从刻板印象的表征特点出发，就信息的承载形式而言，哪种刺激呈现方式会更有助于启动主动控制的加工策略以减少内隐性别刻板印象的表达呢？最后，在现实生活中除了使主动控制在一定的条件下发挥作用以外，更重要的是在当前情况下引发的主动控制能为后续的情况创造控制准备状态(Kleiman et al., 2014)，使得主动控制可以在不同的刻板印象情况下进行控制调整。有研究者曾使用试次-试次控制调整范式(trial to trial control adjustment paradigms; Kleiman et al., 2014)也证明了无关任务(经典Flanker 任务)的控制准备状态对后续刻板印象任务的表达起到了跨领域的控制调整(Kleiman et al., 2014)。所以，我们关心不同冲突水平下的控制准备状态对随后性别刻板印象任务会有着怎样的影响？

新增参考文献：

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), 7351–7356.

Chen, L., Zeng, S, Y., & He, R. (2021). Negative emotions can interfere with the inhibitory effect of proactive control on implicit stereotypes. *North American Academic Research Journal*, 4(4),53–69.

Gonthier, C., Ambrosi, S., & Blaye, A. (2021). Learning-based before intentional cognitive control: developmental evidence for a dissociation between implicit and explicit control. *Journal of Experimental Psychology Learning Memory and Cognition*.

Gross, E.(1968). Plus Ca Change...? The Sexual Structure of Occupations Over Time. *Social Problems* 16, 198-208.

Koenig, A. M., Eagly, A. H. 2014. "Extending role congruity theory of prejudice to men and women with sex-typed mental illnesses." *Basic & Applied Social Psychology*, 36(1), 70-82.

Veen, V. V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI Study. *NeuroImage*, 27(3), 497–504.

**意见 2:** 对于研究中所采用的图片材料和文字语义材料，这些刺激材料可能产生不同结果以及理论需要在问题提出中补充说明；

**回应:** 感谢审稿专家提出的意见。根据审稿专家的宝贵意见，对引言部分进行了补充修改，具体修改如下：

从刻板印象的图式化语义表征特点出发，以具体图像信息作为考察刻板印象认知加工的目标刺激，其研究结果对刻板印象加工机制的解释不可避免存在一定的局限(王沛等, 2010)。更为重要的是，就信息的承载形式而言，知觉者在社会生活中常借助以词语或句子为形式的语义信息来进行日常交流，若是以被赋予一定社会属性的图片信息作为目标刺激，不论所选取的信息与某一群体的联系有多紧密(例如，男性-工作场景，女性-家庭场景等)，都无法完全表现出对某一群体的刻板印象概念化的抽象表征。例如，Chen & Zeng (2021) 的研究中，就采用图片分类确认方式，得出在消极情绪下主动控制模式更多的出现在高冲突水平中，但是语义加工是一种自上而下的加工，而且更加抽象，因此不同信息的呈现方式是否对刻板印象的激活和表达，以及这两种表征方式是否有助于启动主动控制策略，有所不同需要进一步研究。

**参考文献:**

Chen, L., Zeng, S, Y., & He, R. (2021). Negative emotions can interfere with the inhibitory effect of proactive control on implicit stereotypes. *North American Academic Research Journal*, 4(4),53–69.

**意见 3:** 研究 2 和研究 3 之间的逻辑需要更清晰地说明，从实验材料和范式来看，这两者的递进关系不明显；

**回应：**感谢审稿专家提出的意见。根据审稿专家的宝贵意见，做出如下说明：

实验 2 是词语-分类任务，呈现一张男性或者女性面孔，紧接着出现不同的代表家庭或者工作的词语，让被试作出一致或者不一致的反应。而实验三是在实验 2 的基础上，利用改进的试次-试次控制调整范式(trial to trial control adjustment paradigms; Kleiman et al., 2014)，每个完整试次都由两个连续的试次组成，共包含两个不同的行为反应，前一个试次是使用实验 2 中使用到的词语-分类任务，为了验证不同冲突水平对后续加工任务的影响，接着第二个任务是性别 Flanker 任务。在这个任务中，被试会在屏幕中央看到一个男性名字(如，李军)或者一个女性名字(如，刘婷)，同时每个名字的左面和右面会出现与性别一致/不一致的刻板特质词，被试通过按键对名字的性别做出判断。

实验 3 的主要目的是为了证明高冲突水平的任务引发的控制加工对后续不同冲突类型任务的加工存在跨情境调整的一致性(Egner, 2008; Freitas et al., 2007)，即高冲突水平下产生的认知控制加工能够在随后的任务中被持续保持，不受任务类型变化的影响。

**意见 4：**对于文中一些术语的表达，如冲突条件，刻板不一致条件等有待区分，此外，“内隐”性别刻板印象的内隐如何体现也有待交代；

**回应：**感谢审稿专家提出的意见。根据审稿专家的宝贵意见，对引言部分进行了补充修改并做出如下说明：

冲突条件在文中指的是刻板印象不一致条件，冲突条件包括高冲突、基线以及低冲突三种水平。

内隐性别刻板印象的体现：首先，在本研究中，三个实验均是采用分类-确认任务，让被试对家庭或者工作图片进行分类，并没有体现明显的性别刻板印象任务。其次，在 Amodio 等(2018)的研究中，三个实验也均采用相同的实验流程，在不同的冲突水平下验证主动控制是否会消除内隐种族偏见和种族刻板印象；最后，在 Chen & Zeng (2021)的研究中，也采用图片-分类确认任务，启动刺激为男性或者女性图片，目标刺激为笔或者注射器的图片，被试对笔和注射器分类，验证了在消极情绪状态下，主动控制对内隐性别刻板印象的影响。这两个研究均是采用相同的实验流程，验证主动控制对内隐刻板印象的影响。

**意见 5：**总讨论部分略显笼统，对于本研究的主要理论创新和贡献有待丰富和拓展，建议采用小标题更有层次地论述。

回应：非常感谢审稿专家对本文耐心细致的审阅，以及所提的建设性意见。通过对总讨论的重新梳理。本次修改补充的内容均用红色在原文中标识。具体修改内容如下：

## 5 总讨论

不同冲突水平的任务背景和不同条件诱发的认知控制对性别刻板印象表达有着不一样的效果。实验 1 的结果表明，无论是何种冲突试次的比例水平，图片分类任务均能抑制性别刻板印象的表达；而对于实验 2 的词语分类任务而言，相较于基线水平，高冲突水平下性别刻板印象的表达受到了抑制，低冲突水平下性别刻板印象的表达得到了激活；实验 3 的结果则进一步表明高冲突水平下产生的认知控制加工能够在随后的任务中被持续保持，不受任务类型变化的影响。这些发现暗示更大的冲突试次比例可能会对控制刻板印象的表达更有效，而控制的有效性和刻板印象的具体呈现形式有关，个体可以敏感觉察出冲突大小的精细变化，且高冲突任务引发的冲突适应为后续任务提供了控制准备状态(Bukowski et al., 2019; Schmidt, 2019)。

### 5.1 任务冲突比例对性别刻板印象表达的影响

经典的无社会意义的认知冲突实验研究已表明，冲突试次占比高的实验条件可诱发一种主动控制模式(Gratton et al., 1992)。实验 2 和实验 3 使用加工分离程序的分析结果也发现，相对于低冲突条件，高冲突条件下的控制加工值(PDP-C)更大(基线时的控制估计为中间值)。这间接说明了高冲突水平代表了更多的控制性加工，而不是代表了更多性别刻板印象的反应性激活。从常识上来看，更强的干扰环境导致判断准确性的提高这一发现可能有违直觉，但确与主动控制的机制以及过去的研究有着直接的相似性(Appelbaum et al., 2014)。通过进一步对自动加工值(PDP-A)的数据处理，我们发现，主动控制和反应控制可以协同作用于内隐偏见的行为表达，而不是偏见的来源(如自动激活的刻板印象)。另有研究者发现，主动控制与反应控制并非独立存在，在知觉到冲突有所变化时，个体可同时运用两种认知控制策略以形成正确的反应(Bugg & Braver, 2016)，实验结果也充分论证了主动控制和反应控制能根据情况需求产生一定互补性的观点。

### 5.2 任务冲突比例下，刻板印象的抑制表达有其限制性

除了关注冲突水平引发的主动控制和反应性加工特点，同样值得关注的是，不同冲突水平的实验处理在以图片为基础的加工或图片文字混合为基础的加工上存有差异。从实验 1 和实验 2 的结果来说，不同类型的目标刺激材料在高干扰环境下所引发的主动控制和反应性控制加工并不是完全一致的，进一步为冲突水平对信息呈现形式的敏感性存在差异提供了有效证据。这样的发现同刻板印象的不同呈现形式有着很大的相似性。

从实验 1 的分析结果上来看,不同试次比例的冲突任务并没有使个体的认知加工出现明显的差异,出现这种结果的原因可能与实验材料的外在表现维度及内在加工形式有关。首先,实验 1 中的启动刺激和目标刺激均为彩色高清图片,具有很高的生态效度,但这种形式的图片也相应的具有易辨识、易记忆、易加工的特性(Yuan et al., 2019),并且关于文字和图像处理区别的理论观点认为:图像更倾向于自下而上的具体样例形式加工,而文字加工则包含了比图片加工更多的抽象形式加工(Yuan et al., 2019),图像在认知层面减少了自上而下的转换资源,可直接依据具体情境来做出判断(Sherman, 1996),因此可能会出现处理条件对实验结果没有影响的结果,即图片在不同冲突水平下的加工速度和准确性并无优劣之分。实验 2 将图片更换为词语作为目标刺激的呈现形式,结果表明高冲突条件下的性别图片呈现时,抽象文字类型的目标刺激引发的主动控制对性别角色刻板印象的表达有一定的抑制作用。从而验证了这样一个假设,即相对于弱反应干扰,强反应干扰的环境会引起更多的主动控制,进而减少内隐偏见的表达(Amodio et al., 2018)。结果正如预期的那样,相对于基线条件,低冲突条件诱发了一种更夸大的内隐偏见模式,而在高冲突条件下,这种内隐偏见的模式被削弱。

### 5.3 任务冲突比例下,刻板印象的抑制表达具有领域适用性的特点

实验 3 在实验 2 的结论基础上,利用改进的试次-试次控制调整范式证明先前高冲突比例的性别刻板印象分类任务与后续不同类型的性别 Flanker 任务有着相似的认知控制模式。最近一些研究也充分证明冲突状态能够为后续任务提供控制准备,进而引发认知控制动态加工中更多的冲突适应(Kleiman et al., 2014; Bukowski et al., 2019; 张孟可等, 2021)。Bukowski 等(2019)为了评估认知灵活性,使用了包含社会类别的任务转换程序,在实验 1 中,被试按性别或年龄分类,在实验 2 和实验 3 中,被试按性别或社会角色分类,同样都发现被试的认知灵活性主要表现在引起更多认知冲突的任务情况下,并且认知灵活性在不同的任务冲突比例中体现为控制策略的灵活调整。因此,被试在高冲突背景下的性别 Flanker 任务中,更易引发主动性控制来抑制 Flanker 刺激诱发的干扰,将注意力集中在位于中央位置的目标刺激上。低冲突背景下则更容易受到反应性加工的诱导,从整个刺激界面中提取信息并做出反应。

综合而言,以往大量的研究关注了认知控制可灵活转变的特性,但是都局限在相同的冲突任务背景内进行比较,也常忽略信息表征的特殊表现形式和社会刺激的不同类型之间的控制准备作用。由于我们日常各类行为活动都是在不同比例的冲突社会环境中进行的,因而考虑到任务的冲突比例和社会属性是很有必要的。并且从现实生活来看,形成传统规范的性别角色观念并不容易新的社会规范所打破,以群体为基础的性别刻板印象及偏见可能会继续抵制做出改变。在这样的情况下,不主张关注性别与刻板信息之间的联结,而是主张关注



任务中的目标信息，可能为削弱性别刻板印象的表达提供了一种有效的方式。不同的任务冲突比例引发的认知控制对内隐偏见的表达有一定的调节作用。不同呈现形式的刺激信息和具社会意义的转换范式，证明了认知层面的抑制性控制可在特定的呈现形式和跨任务之间减少偏见反应的出现，使得内隐偏见的减少和消除不仅仅局限于刻板印象出现后再进行反应性的加工。

尽管本研究有一些重要的发现，但也存在一些不足。首先，认知控制是目标导向行为的基础，个体是如何根据不同的冲突任务感知对当前目标进行认知控制调整的不得而知，由练习效应而诱发的比例效应和基于目标而引发的比例效应有何区别。这些问题仅借助行为分析并不能使得到有效的回答，也不能保证实验结果的充分性。所以观察不同冲突比例的任务过程中的脑电水平，识别冲突监测和控制调整的脑区差异，有助于我们从客观的角度深入探究冲突水平背后的认知神经机制，为我们在不同的冲突环境下进行有效的认知控制提供理论基础。其次，冲突适应效应和比例效应是认知控制的两种表现形式，冲突适应效应强调个体对先前试次冲突的监测导致在当前试次中更好地控制了冲突，而比例效应则更为关注同一冲突任务类型内监测到不同试次冲突水平的差异引发的冲突适应现象。该研究并没有涉及到两种效应的对比分析，也未深入思考两种效应引发的认知控制有何异同之处。

另外，Yang 等(2017)利用冲突适应效应的减少来检验认知控制是否可以在不同通道的冲突任务中转换，结果表明不同冲突任务下出现的认知控制加工存在着通道特异性。这为我们探究认知控制对抑制各种刻板印象的表达提供了新的思路，考虑不同任务情境和任务难度在不同的加工通道中如何进行认知控制转换，转换成功的边界条件等研究问题。最后，以关注目标为主，试图摆脱干扰信息的认知控制能力固然对改变刻板印象有着重要意义，但无时无刻充斥在我们的生活中的性别刻板印象，并不能如同实验般有着固定的刺激类型、以认知基础的实验任务以及设置好的冲突情境水平，而是常在不同的场合以不同的刺激形式组合而成，所以如何克服真实情境中的干扰信息对性别刻板印象的认知控制的影响仍需进一步探索。

**意见 6:** 参考文献的引用有待进一步符合《心理学报》的最新规范。

**回应:** 感谢审稿专家提出的意见。根据审稿专家的宝贵意见，对参考文献的引用进行了梳理修改，本次修改补充的内容均用红色字体对原文相应内容进行标识。

启动图片的材料参考 Croft(2016)研究 1，结合已有图库选取男女图片各 10 张，用 Photoshop7.0 软件对选取照片进行统一标准化处理，使所有照片均为相同亮度的灰色背景，

大小均为 420×380 像素(Amodio & Swencionis, 2018)。

由于个体本身的初始情绪与性别角色的观念水平会对刻板印象的认知加工产生影响。使用邱林, 郑雪和王雁飞(2008)修订的《PANAS 情绪自评量表》来测量被试的情绪水平, 该情绪量表采用 5 点计分的方式, 包括积极情绪自评(9 个积极形容词)和消极情绪(9 个消极形容词)自评两部分, 每种情绪类型的均分越高, 表明情绪越强烈。结合中国妇女社会地位调查(CWSSS)和中国综合社会调查(CGSS)中的性别观念的题项形成《性别角色观念量表》, 将其区分为传统主义取向和现代主义取向(Rajadhyaksha et al., 2015)。该实验中, 《性别角色观念量表》的信度系数为 0.83, 结构效度为 0.67, 60 名被试在该量表中的得分为  $M=3.80$ ,  $SD=0.51$ 。之后采用改进的武器识别范式(Weapons Identification Task, Payne, 2001)和操纵任务干扰的方法(Appelbaum et al., 2014)来操控实验流程。

刻板印象表征的联结模型指出, 分布式的循环网络善于对一些图像或模式干扰进行识别, 基于特定任务出现的循环联结可以使个体从本质上过滤出各种干扰(Van et al., 2003), 所以不强调启动图片/干扰刺激与目标图片的联结判断, 使得个体更易获得只关注目标图片的加工模式, 其相应的判断会较少受到同一冲突类型但存在冲突水平差异的影响。

我们使用加工分离程序(Process Dissociation Procedure, PDP)进一步验证我们的假设(Jacoby, 1991; Jacoby et al., 1999)。

高冲突水平和基线水平分类任务之后的性别 Flanker 任务的反应时高于低冲突分类任务之后的性别 Flanker 任务的反应时。这些结果均为内隐偏见任务可能会显著改变冲突监测和控制调整过程运作的环境提供了有力的支撑(Amodio et al., 2008; Freitas et al., 2007)。

.....

### 审稿人 3 意见:

当今社会依旧存在传统的“男主外女主内”的刻板观念和男女社会行为上的差异。本文以不同冲突水平的认知控制为切入点, 探讨了其对于性别刻板印象的影响, 较为新颖。但是, 研究也有一些需要克服和澄清的地方。意见 1: 多处语言表述可斟酌修改, 如“但鉴于社会中仍有许多力量不断地促进和维持人们思想中的内隐偏见, 通过改变个体。”这句话, 前后两句话实际上没有关联; 如“开展了比较减少内隐种族偏见表达的 17 项措施效果的实验, 但结果仅 8 项有效地减少了”, “而不是偏见的一个具体例子”, “间接地来抑制了行为中偏见因素的表达”等。意见 2: 研究假设直接提出对于各种条件下的正确率和反应时, 以及性别刻板印象的表达是否被抑制的猜想, 但在前面的文献证据中并未提及假设的一些依



据，可补充支撑的文献。意见 3：实验一的结果报告中，主效应和交互作用均不显著的情况下，有的加了  $p > 0.05$ ，有的没有，请统一。讨论中“即使三种冲突水平下的正确率并没有显著的差异，但正确率和控制的估算值都是相对较高的”，但作者并未在结果中给出相应的值，我无法得知。意见 4：实验三中，首先做了冲突水平×目标类别×特质词类别的三因素重复测量方差分析，发现三者交互作用显著，但并未直接做此交互作用的简单效应分析并进行解释说明，这是为什么？意见 5：注意流程图和图表的规范化，如流程图上的中文标注不应该是单独插入的，而是一张整图；图 6 最右边的差异线应该将两个组包含在内。注意结果表述的严谨性，“存在比较显著的差异 ( $p < 0.001$ )”， $p < 0.001$  的差异不应该称作比较显著。意见 6：最重要的一点是，文章提出了不同冲突水平的认知控制对性别刻板印象的影响，发现低冲突水平易加强性别刻板印象的表达，高冲突水平下产生的认知控制加工能够在随后不同的任务中得到保持。本研究最有价值的点（如实验能为性别刻板印象带来什么样的改变）并未在讨论中阐述，即实验为进一步消除性别偏见可以提供什么样的理论支撑或是有效方法？

意见 1：多处语言表述可斟酌修改，如“但鉴于社会中仍有许多力量不断地促进和维持人们思想中的内隐偏见，通过改变个体。”这句话，前后两句话实际上没有关联；如“。。。开展了比较减少内隐种族偏见表达的 17 项措施效果的实验，但结果仅 8 项有效地减少了”，“而不是偏见的一个具体例子”，“间接地来抑制了行为中偏见因素的表达”等。

回应：感谢审稿专家提出的意见。根据审稿专家的宝贵意见，对文中语句的进行了斟酌修改，本次修改补充的内容均用红色字体对原文相应内容进行标识。具体修改内容如下：

传统的性别刻板印象对现代女性职业和家庭之间的协调有着很大的困扰，而男性也承担着来自家庭经济主力的巨大压力。这种刻板化的家庭角色分工对性别平等和家庭和谐往往会产生消极的影响(Cundiff & Vescio, 2016)。虽然在社会层面，通过设置陪产假、育婴假、利用社会规范等干预措施来缓解男性和女性在职业和家庭中的刻板印象。但即便如此，扎根于人们思想中的内隐偏见依然难以改变，并且可能会被自动激活，所以通过改变个体在实际任务中的认知控制来抑制性别刻板印象的表达，可能会为进一步消除这种性别偏见提供了有效方法。

认知控制指的是人们能根据目标或任务做出相应的反应，遇到分心物能保存目标、抑制习惯性反应或冲动行为的能力(Miller & Cohen, 2001)，是目标导向行为的基础，在实验室条件下会采用冲突任务来研究认知控制(Gonthier & Ambrosi, 2021)。冲突监控理论(conflict

monitoring theory; Veen & Carter, 2005)是最常见的对冲突任务的解释。该理论认为,冲突受到自上而下认知控制的调节,当先前的试次是不一致试次会引发调整机制,认知控制水平提高,但当先前试次为一致试次,则无需调整,即个体在先前试次的冲突检测导致在当前试次更好地控制了冲突(Botvinick et al., 2001; Botvinick et al., 1999)。大量研究表明认知控制分为两类,对快速变化的外部条件作出即时行动一般属于认知控制过程中的反应性控制,对快速变化的外部条件做出更为持久的战略优化行动一般属于认知控制过程中的主动控制(Funes et al., 2010; Braver, Paxton et al., 2009)。例如,一个人在下班后总是在十字路口左转,但听说那条路上堵车,他可能在到达十字路口之前提醒自己右转,这就是主动控制;但他也可能采用反应控制模式,即到达十字路口后,遇见堵车,然后右转以避免拥挤的交通。此时,反应控制比主动控制需要的认知资源少,但更容易出错。

迄今为止,几乎所有关于控制刻板印象或偏见的研究都集中在反应控制上。反应控制是在刻板印象的被检测和发现后,对其进行反应性的抑制(Mark & Muraven, 2008)。反刻板样例暴露法(Ramasubramanian, 2015)、生动反刻板印象情景故事法(Burnet et al., 2017)、虚假 IAT 结果反馈(Cvencek et al., 2010)等方法均属于反应控制的干预训练法。以往的研究表明,这些方法能够暂时有效地削弱刻板印象带来的影响(Lai et al., 2016)。然而,尽管反应性控制为克服偏见提供了一个重要的机制,它只在偏见出现或激活之后才开始发挥作用,因此容易受到社会中各种情境因素的不断强化以及认知资源的影响,从而破坏它的有效性,出现干预效果消退的情况(Lambert et al., 2003; Calvin, 2016)。

不同于反应性控制,主动控制是一种完全不同的干预概念。作为一种目标相关的控制方式,它针对的是一个人的预期反应,而不是偏见的来源。主动控制采用了操纵任务干扰的方法,它可以在激活人们头脑中关于刻板印象的相关信息之前,有效地将刻板印象的关注点迁移至任务目标本身,通过对任务困难的体验或对一个人的预期行为增加的动机来限制内隐偏见在行为中表达的可能性(Amodio et al., 2018)。已有研究表明,在种族分类任务中观察到的刻板印象偏见的反应主要是由冲突驱动的,即高冲突增强了被试对任务目标的关注,可能导致积极主动的控制模式(Bartholow & Dickter, 2008; Dickter & Bartholow, 2010),但这一研究并未直接验证主动控制对种族偏见的干预。在一项最新研究中(Amodio et al., 2018),研究者提出了主动控制模型(Proactive Model of Control)(图 1),采用了武器识别范式(Payne, 2001)、偏见启动任务与刻板印象启动任务考察了主动控制对内隐偏见和内隐刻板印象表达的抑制作用,首次证实了相比基线组,个体在高冲突水平下会采用主动控制的加工模式,抑制偏见或者刻板印象的表达,在低冲突水平下则会更明显地表现出内隐种族偏见及内隐种族刻板印

象。然而，以上研究中，主要是针对种族相关的偏见行为。然而，在现实生活中，与性别相关的刻板印象与每一个个体都息息相关，尤其体现在男性和女性更适合从事哪一类职业的固定观念上，而这种观念会导致职业性别产生隔离(Gross, 1968)，从而对不同行业的就业产生消极影响(Koenig & Eagly, 2014)。遗憾的是，主动控制对性别刻板印象的影响并未引起研究者的广泛关注，所以，这一认知加工方式是否能同样有效地抑制性别刻板印象还有待探究。

参考文献：

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), 7351–7356.

Chen, L., Zeng, S, Y., & He, R. (2021). Negative emotions can interfere with the inhibitory effect of proactive control on implicit stereotypes. *North American Academic Research Journal*, 4(4),53–69.

Gonthier, C., Ambrosi, S., & Blaye, A. (2021). Learning-based before intentional cognitive control: developmental evidence for a dissociation between implicit and explicit control. *Journal of Experimental Psychology Learning Memory and Cognition*.

Gross, E.(1968). Plus Ca Change...? The Sexual Structure of Occupations Over Time. *Social Problems* 16, 198-208.

Koenig, A. M., Eagly, A. H. 2014. “Extending role congruity theory of prejudice to men and women with sex-typed mental illnesses.” *Basic & Applied Social Psychology*, 36(1), 70-82.

Veen, V. V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI Study. *NeuroImage*, 27(3), 497–504.

**意见 2：**研究假设直接提出对于各种条件下的正确率和反应时，以及性别刻板印象的表达是否被抑制的猜想，但在前面的文献证据中并未提及假设的一些依据，可补充支撑的文献。

**回应：**感谢审稿专家提出的意见。根据审稿专家的宝贵意见，对文献进行了补充，对引言部分进行了修改，具体修改内容如下：

迄今为止，几乎所有关于控制刻板印象或偏见的研究都集中在反应控制上。反应控制是在刻板印象的被检测和发现后，对其进行反应性的抑制(Mark & Muraven, 2008)。反刻板样例暴露法 (Ramasubramanian, 2015)、生动反刻板印象情景故事法 (Burnet al., 2017)、虚假 IAT 结果反馈 (Cvencek et al., 2010) 等方法均属于反应控制的干预训练法。以往的研究表明，这些方法能够暂时有效地削弱刻板印象带来的影响(Lai et al., 2016)。然而，尽管反

应性控制为克服偏见提供了一个重要的机制，它只在偏见出现或激活之后才开始发挥作用，因此容易受到社会中各种情境因素的不断强化以及认知资源的影响，从而破坏它的有效性，出现干预效果消退的情况(Lambert et al., 2003; Calvin, 2016)。

不同于反应性控制，主动控制是一种完全不同的干预概念。作为一种目标相关的控制方式，它针对的是一个人的预期反应，而不是偏见的来源。主动控制采用了操纵任务干扰的方法，它可以在激活人们头脑中关于刻板印象的相关信息之前，有效地将刻板印象的关注点迁移至任务目标本身，通过对任务困难的体验或对一个人的预期行为增加的动机来限制内隐偏见在行为中表达的可能性(Amodio et al., 2018)。已有研究表明，在种族分类任务中观察到的刻板印象偏见的反应主要是由冲突驱动的，即高冲突增强了被试对任务目标的关注，可能导致积极主动的控制模式(Bartholow & Dickter, 2008; Dickter & Bartholow, 2010)，但这一研究并未直接验证主动控制对种族偏见的干预。在一项最新研究中(Amodio et al., 2018)，研究者提出了主动控制模型(Proactive Model of Control)(图 1)，采用了武器识别范式(Payne, 2001)、偏见启动任务与刻板印象启动任务考察了主动控制对内隐偏见和内隐刻板印象表达的抑制作用，首次证实了相比基线组，个体在高冲突水平下会采用主动控制的加工模式，抑制偏见或者刻板印象的表达，在低冲突水平下则会更明显地表现出内隐种族偏见及内隐种族刻板印象。然而，以上研究中，主要是针对种族相关的偏见行为。然而，在现实生活中，与性别相关的刻板印象与每一个个体都息息相关，尤其体现在男性和女性更适合从事哪一类职业的固定观念上，而这种观念会导致职业性别产生隔离(Gross, 1968)，从而对不同行业的就业产生消极影响(Koenig & Eagly, 2014)。遗憾的是，主动控制对性别刻板印象的影响并未引起研究者的广泛关注，所以，这一认知加工方式是否能同样有效地抑制性别刻板印象还有待探究。

从刻板印象的图式化语义表征特点出发，以具体图像信息作为考察刻板印象认知加工的目标刺激，其研究结果对刻板印象加工机制的解释不可避免存在一定的局限(王沛等, 2010)。更为重要的是，就信息的承载形式而言，知觉者在社会生活中常借助以词语或句子为形式的语义信息来进行日常交流，若是以被赋予一定社会属性的图片信息作为目标刺激，不论所选取的信息与某一群体的联系有多紧密(例如，男性-工作场景，女性-家庭场景等)，都无法完全表现出对某一群体刻板印象概念化的抽象表征。例如，Chen & Zeng (2021)的研究中，采用图片分类确认方式，得出在消极情绪下主动控制模式更多的出现在高冲突水平中，但是语义加工是一种自上而下的加工，而且更加抽象，因此不同信息的呈现方式是否对刻板印象的激活和表达，以及这两种表征方式是否有助于启动主动控制策略，有所不同需要进一步研究。

随着刻板印象研究的深入,关于不同冲突比例水平下的认知控制的基础研究已经取得了相当大的进展(李政汉等, 2018),但在不同的冲突比例水平下,认知控制对图片刺激所诱发的性别刻板印象表达会有着怎样的影响呢?此外,从刻板印象的表征特点出发,就信息的承载形式而言,哪种刺激呈现方式会更有助于启动主动控制的加工策略以减少内隐性别刻板印象的表达呢?最后,在现实生活中除了使主动控制在一定的条件下发挥作用以外,更重要的是在当前情况下引发的主动控制能为后续的情况创造控制准备状态(Kleiman et al., 2014),使得主动控制可以在不同的刻板印象情况下进行控制调整。有研究者曾使用试次-试次控制调整范式(trial to trial control adjustment paradigms; Kleiman et al., 2014)也证明了无关任务(经典 Flanker 任务)的控制准备状态对后续刻板印象任务的表达起到了跨领域的控制调整(Kleiman et al., 2014)。所以,我们关心不同冲突水平下的控制准备状态对随后性别刻板印象任务会有着怎样的影响?

目前,较少有研究直接涉及到不同冲突水平的认知控制对性别刻板印象的影响。针对上述问题,本研究假设:(1)在图片分类-确认实验任务中,与基线水平相比,被试在高冲突水平下驱动更强的主动控制调整,性别刻板印象的表达更容易被抑制。(2)在词语分类-确认实验任务中,与基线水平相比,被试在高冲突水平下驱动更强的主动控制调整,性别刻板印象的表达更容易被抑制。(3)在先前不同冲突水平的词语分类任务后,高冲突水平下的反应更能抑制后续刻板 Flanker 任务性别刻板印象的表达,驱动更强的主动控制。

意见 3: 实验一的结果报告中,主效应和交互作用均不显著的情况下,有的加了  $p>0.05$ ,有的没有,请统一。讨论中“即使三种冲突水平下的正确率并没有显著的差异,但正确率和控制的估算值都是相对较高的”,但作者并未在结果中给出相应的值,我无法得知。

回应:感谢审稿专家提出的意见。根据审稿专家的宝贵意见,本次修改补充的内容均用红色字体对原文相应内容进行标识。具体修改内容如下:

将性别角色观念均分作为协变量,在冲突水平上对被试反应的平均正确率和反应时进行重复测量方差分析,结果(如表 1)并未出现任何显著的主效应和交互作用( $p>0.05$ )。

对每种冲突条件下刺激类型的平均正确率进行 3(冲突水平:高冲突,基线,低冲突) $\times$  2(启动刺激:男性图片,女性图片) $\times$  2(目标刺激:工作图片,家庭图片)的重复测量方差分析,结果显示:启动刺激和目标刺激之间的交互作用显著,  $F(1, 59)=4.91, p=0.031, \eta_p^2=0.08$ 。其他的主效应和交互作用均不显著( $p>0.05$ )。

对每种冲突条件下刺激类型的反应时均值进行 3(冲突水平：高冲突水平，基线水平，低冲突水平)×2(启动刺激：男性图片，女性图片)×2(目标刺激：工作图片，家庭图片)的重复测量方差分析，结果并未显示出任何显著的效应( $p>0.05$ )。

表 1 不同冲突水平下和目标刺激下目标分类的正确率及反应时

冲突水平	启动刺激× 目标刺激	正确率	反应时	总正确率	总反应时
		<i>M(SE)</i>	<i>M(SE)</i>	<i>M(SE)</i>	<i>M(SE)</i>
高	男-工作	0.96(0.01)	452.89 (14.63)	0.96 (0.01)	451.74 (12.90)
	男-家庭	0.94(0.01)	451.04 (13.24)		
	女-工作	0.96(0.01)	451.97 (13.54)		
	女-家庭	0.96(0.01)	451.05 (14.07)		
基线	男-工作	0.96(0.01)	456.01 (13.58)	0.95 (0.01)	455.46 (12.38)
	男-家庭	0.94(0.01)	461.46 (13.44)		
	女-工作	0.95(0.01)	456.98 (13.31)		
	女-家庭	0.94(0.01)	447.39 (12.37)		
低	男-工作	0.96(0.01)	453.36 (14.49)	0.95 (0.01)	459.34 (13.97)
	男-家庭	0.95(0.01)	459.01 (14.86)		
	女-工作	0.95(0.01)	466.83 (15.21)		
	女-家庭	0.95(0.01)	458.16 (15.20)		

表 2 不同水平下的主动控制与反应控制估算值

	主动控制 <i>M (SE)</i>	反应控制 <i>M (SE)</i>
高冲突	0.92(0.01)	0.79(0.06)
基线	0.90(0.01)	0.49(0.06)
低冲突	0.91(0.01)	0.42(0.06)



**意见 4:** 实验三中，首先做了冲突水平×目标类别×特质词类别的三因素重复测量方差分析，发现三者交互作用显著，但并未直接做此交互作用的简单效应分析并进行解释说明，这是为什么？

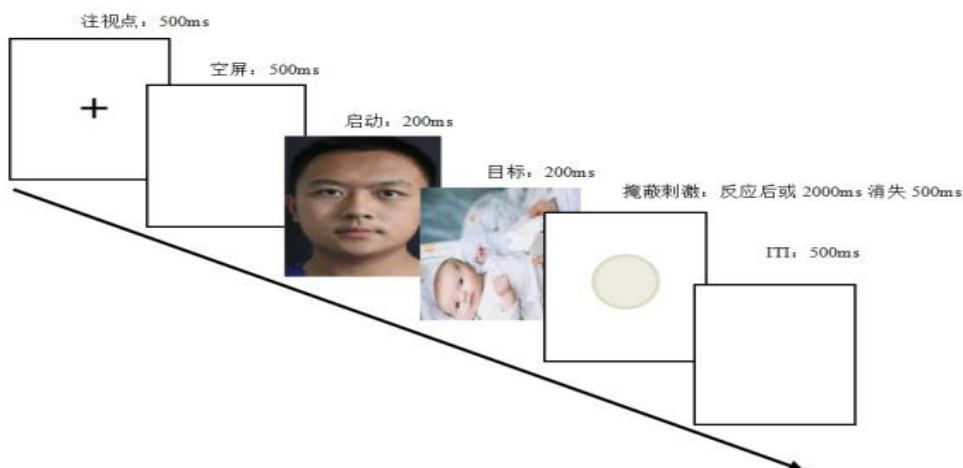
**回应:** 感谢审稿专家提出的意见。根据审稿专家的宝贵意见，做出如下说明：

在本研究中，我们更关注在三种不同的冲突水平下，启动刺激和目标刺激是否产生交互作用。如果启动刺激和目标刺激产生交互作用，则说明存在典型的偏见模式；如果启动刺激和目标刺激不存在交互作用，说明性别刻板印象表达得到抑制。所以，通过分解三阶交互作用，说明在不同冲突水平下，启动刺激和目标刺激的统计结果，这样会更加清楚的显示我们的研究假设以及研究更关注的问题。

**意见 5:** 注意流程图和图表的规范化，如流程图上的中文标注不应该是单独插入的，而是一张整图；图 6 最右边的差异线应该将两个组包含在内。注意结果表述的严谨性，“存在比较显著的差异( $p < 0.001$ )”， $p < 0.001$  的差异不应该称作比较显著。

**回应:** 非常感谢审稿专家提出的问题。我们对文中的图表及结果表达进行了修改完善，具体内容如下，本次修改的文字内容均用红色字体对原文相应内容进行标识。

具体修改内容如下：



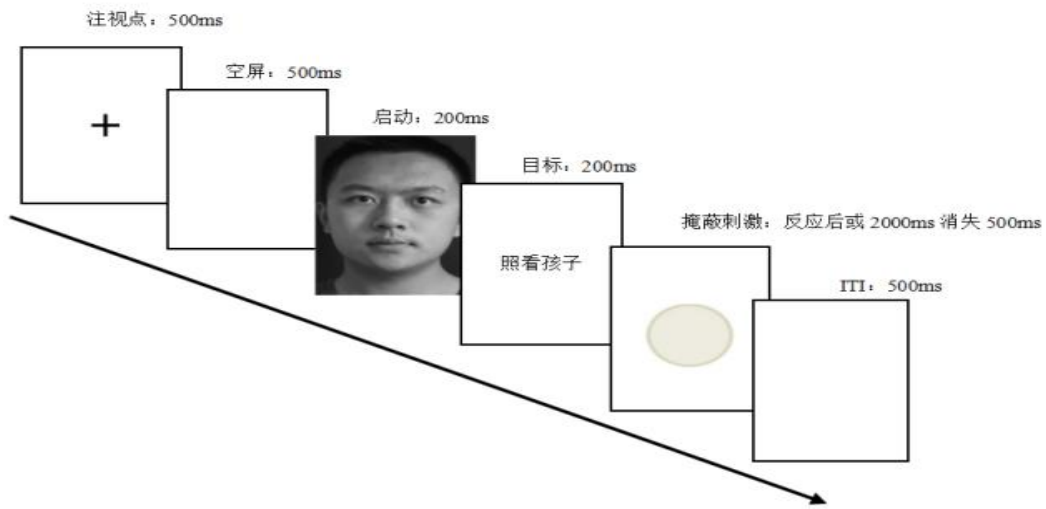


图3 实验2流程图

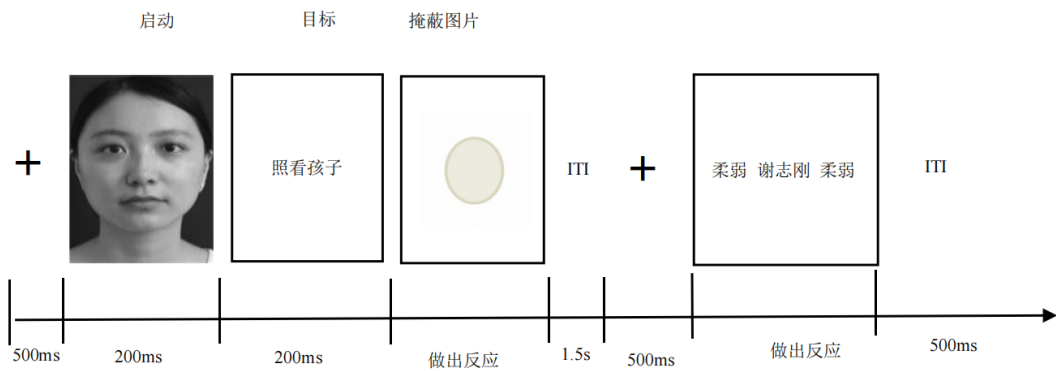


图5 实验3流程图

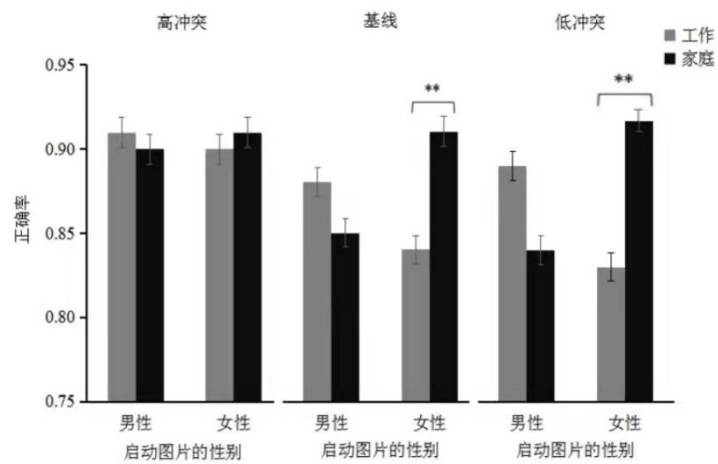


图6 不同冲突水平和启动刺激下目标分类的正确率



启动图片从实验 1 筛选出的性别图片中选取。目标词语的来源参考吴梦玲(2017)的研究,再通过网络调查的方式获取 20 个家务活动和 20 个职业工作的词语,然后选取 40 名被试(其中男性 19 人,  $M=20$ )对词语的典型性进行 5 点评分(从 1 “非常不典型” —5 “非常典型”)。对筛选出的词语做差异性检验,独立样本 t 检验的结果表明,家庭词语与工作词语在典型性上无显著差异( $M_{\text{家庭}}=3.95$ ,  $M_{\text{工作}}=3.97$ ),  $t(38)=-0.288$ ,  $p=0.775$ ; 单样本 t 检验的结果表明,家庭词语和工作词语均与典型性的中间水平存在**极其显著差异**( $p<0.001$ )。最后选取每组词平均得分靠前的词语作为目标词语。

对冲突水平的主效应做进一步的事后多重比较,如表 4 的结果显示:高冲突水平下四种刺激类型的平均反应时( $M_{\text{高冲突}}=525.33$ ,  $SD=15.81$ )与低冲突水平( $M_{\text{低冲突}}=465.52$ ,  $SD=13.22$ )存在**极其显著差异**( $p<0.001$ ); 基线水平( $M_{\text{基线}}=513.87$ ,  $SD=13.10$ )与低冲突水平( $M_{\text{低冲突}}=465.52$ ,  $SD=13.22$ )差异显著( $p<0.001$ ); 不过高冲突水平( $M_{\text{高冲突}}=525.33$ ,  $SD=15.81$ )与基线水平( $M_{\text{基线}}=513.87$ ,  $SD=13.10$ )差异不显著( $p=0.244$ )。被试在高冲突水平上反应时最长,而在低冲突水平上反应时最短。

启动图片的材料参考 Croft(2016)研究 1,结合已有图库选取男女图片各 10 张,用 Photoshop7.0 软件对选取照片进行统一标准化处理,使所有照片均为相同亮度的灰色背景,大小均为  $420 \times 380$  像素(Amodio et al., 2018)。另找 176 名被试作为独立样本,对选取的 20 张图片进行面孔整体印象的 5 点评分(从 1 “非常不好” —5 “非常好”),从中筛选出面孔整体印象位于中间水平的男女图片各 8 张。然后对筛选出的图片进行评分的差异性检验,独立样本 t 检验的结果表明,男性图片与女性图片在整体印象维度上不存在显著差异( $M_{\text{男}}=3.10$ ,  $M_{\text{女}}=3.20$ ),  $t(14)=-1.494$ ,  $p=0.157$ ; 此外,研究表明,人们对不同年龄的男性和女性存在着不同的刻板印象(任娜, 佐斌, 侯飞翔, 汪国驹, 2012),所以本研究将面孔年龄的范围控制在 25~30 岁之间,并且男女图片在面孔年龄上也无显著差异( $M_{\text{男}}=27.95$ ,  $M_{\text{女}}=27.37$ ),  $t(14)=0.492$ ,  $p=0.631$ 。另外,单样本 t 检验的结果表明,男性图片和女性图片与较极端的面孔整体印象水平均有**极其显著差异**( $p<0.001$ )。

**意见 6:** 最重要的一点是,文章提出了不同冲突水平的认知控制对性别刻板印象的影响,发现低冲突水平易加强性别刻板印象的表达,高冲突水平下产生的认知控制加工能够在随后不同的任务中得到保持。本研究最有价值的点(如实验能为性别刻板印象带来什么样的改变)并未在讨论中阐述,即实验为进一步消除性别偏见可以提供什么样的理论支撑或是有效方法?

回应：感谢审稿专家提出的意见。根据审稿专家的宝贵意见，对总讨论进行了补充，本次修改补充的内容均用红色字体对原文相应内容进行标识。具体修改内容如下：

综合而言，以往大量的研究关注了认知控制可灵活转变的特性，但是都局限在相同的冲突任务背景内进行比较，也常忽略信息表征的特殊表现形式和社会刺激的不同类型之间的控制准备作用。由于我们日常各类行为活动都是在不同比例的冲突社会环境中进行的，因而考虑到任务的冲突比例和社会属性是很有必要的。并且从现实生活来看，形成传统规范的性别角色观念并不容易新的社会规范所打破，以群体为基础的性别刻板印象及偏见可能会继续抵制做出改变。在这样的情况下，不主张关注性别与刻板信息之间的联结，而是主张关注任务中的目标信息，可能为削弱性别刻板印象的表达提供了一种有效的方式。不同的任务冲突比例引发的认知控制对内隐偏见的表达有一定的调节作用。不同呈现形式的刺激信息和具社会意义的转换范式，证明了认知层面的抑制性控制可在特定的呈现形式和跨任务之间减少偏见反应的出现，使得内隐偏见的减少和消除不仅仅局限于刻板印象出现后再进行反应性的加工。

---

## 第二轮

审稿人 1 意见：作者针对审稿人的意见逐条进行了认真修改，手稿得到了很大提升，同时也回答了我提出的问题。同意发表。

回应：非常感谢您的肯定！

审稿人 2 意见：论文《基于不同冲突水平的认知控制对性别刻板印象的影响》根据修改意见，作者进行了认真修改，总体质量有了较大提升。建议进一步将语言表达和格式等进一步规范。

回应：非常感谢您的肯定！我们已进行了全文检查，对文中的语言表达和格式进行了修改。

审稿人 3 意见：

基于不同冲突水平的认知控制对性别刻板印象的影响”这个题目是有问题的，到底对刻板印象的什么产生了影响？刻板印象激活还是刻板印象表达？“基于”更是多余。问题提出

中的术语使用比较随意、混乱，如刻板印象与偏见混用，反应抑制和主动抑制的突兀出现等等。话题也比较随意的切换。不清楚到底要研究什么，是主动控制对性别刻板印象的影响还是？一会在说认知控制，一会又在说主动控制。在问题提出甚至是全文中，到底在阐述和研究主动控制还是认知控制？如果是主动控制，有何理由研究？因为没有对比主动控制和反应控制的差异，更谈不上认知控制这个更大的概念。是在研究刻板印象激活还是在研究刻板印象表达？而在实验研究部分并没有很好地操纵与讨论认知控制策略的问题。而且研究者所罗列的假设是无法从问题提出中得来的。也就是说，研究者并没有有效地提出研究怎么样的关键性科学问题。在具体的实验设计中自变量操纵、潜在变量控制、统计方法等诸多环节存在很大的纰漏甚至错误，缺乏清晰的统计分析的逻辑思路。例如：请报告 G-power 对于被试人数的预估计情况以及选用具体人数的理由。实验 3 同实验 1 和实验 2 之间有何必然的联系？在前面的实验的基础上要解决什么问题？因为实验 3 的范式和刺激结构完全同前面的实验不同。实验 1-3 做的是刻板印象激活效应而非刻板印象的表达，因为不存在表达的任务。这一点恰恰与研究者在问题提出与讨论等部分提到的自己研究主题不符。实验 1 中启动刺激仅各自用了 2 张男性-女性照片，这里练习（熟悉）效应对于实验结果的影响是巨大的，如何排除？选取了个 6 张照片，与启动刺激进行配对后在每种刺激条件下仅有 12 张，加上不断重复，这样小而熟悉的刺激集所得到的结果如何确保统计可靠性？在实验 1 中如何控制个体本身的初始情绪与性别角色的观念水平这两个研究者人为地潜在变量？为什么没有数据统计的报告？仅仅把性别角色的观念水平作为协变量进行控制是否足够？在后面的两个实验中为什么对这些控制变量的控制变得更弱？实验 1 中联系正确率达到 70%即为有效数据有何依据，为什么不设置到 90%甚至更高水平？有何依据？实验 1 之所以缺乏显著的结果，一个重要的潜在污染可能是图表的高度可以识别性。不同水平下的主动控制与反应控制估算值是如何得出的，这样的数据就可以说明不同认知控制的水平了吗？为什么在实验 2 出现了各 4 张图片的启动刺激？而实验 1 则是各 2 张？按理说图片加工条件需要更多的启动图片以降低熟悉效应。实验 2 为什么仅仅试图控制性别角色观念水平，而且与实验 1 的控制方法不同？实验 2 在对冲突水平正确率的主效应做进一步的事后多重比较的结果存疑：0.94、0.95、0.96 到底有何差异一定会到显著性水平？那么“但低冲突水平下四种刺激类型的平均正确率(M 低冲突=0.94, SE=0.01)与基线水平差异不显著(M 基线=0.95, SE=0.01, p=0.228)”这个结果又做何解释？实验 2 三阶交互作用的数据量极低，进一步进行统计分析的结果没有什么实质性意义。讨论部分结果阐述与解释混杂在一起，有的解释与结论完全与研究结果无关，难以让人心信服。

意见 1：基于不同冲突水平的认知控制对性别刻板印象的影响”这个题目是有问题的，到底对刻板印象的什么产生了影响？刻板印象激活还是刻板印象表达？“基于”更是多余。

回应：非常感谢审稿专家关于题目的建设性意见。首先，刻板印象的表达(expression of stereotypes),是指间接测量方法所观测到的刻板印象的内隐效应及其程度(Huntsinger et al., 2009)。刻板印象激活(stereotype activation)是指知觉者通过相关线索在头脑中获得(accessibility)靶子所在群体的刻板印象(如特质构念, 行为特征等)(Kunda & Spencer, 2003; Krieglmeier & Sherman, 2012)。因此，刻板印象激活是刻板印象表达的充分而非必要条件，即刻板印象激活不一定会引起个体进行刻板印象的表达，但刻板印象的表达是以刻板印象激活为前提。其次，Amodio 等人(2018)提出，在以往大部分研究中，反应性控制是当检测到偏见后所实施的认知加工方式，其作用是纠正激活的偏见对目标反应的影响，而主动控制可以在激活人们头脑中关于刻板印象的相关信息之前，有效地将刻板印象的关注点迁移至任务目标本身，通过对任务困难的体验或对一个人的预期行为增加的动机来限制内隐刻板印象在行为中表达的可能性，而不仅仅在刻板印象出现或激活之后才开始发挥作用。最后，从实验任务上来看，我们采用的是分类-确认范式而非启动-靶子一致性判断任务，被试需要判断靶刺激的类别(如，家务图片或工作图片)，这一行为结果也是对刻板印象表达的反映但却不能有效地反映刻板印象的激活。此外，冲突检测理论提出，对冲突信息进行加工包含着两个不同的成份：①评价成份(即冲突监测)，②执行成份(即执行控制)。不一致信息在认知过程中所产生的冲突强度是调节认知加工的基础，用于唤起与任务相关的知觉负载。因此，在高冲突背景下，冲突监测会持续发生，执行控制的水平也会适当提高，这时刻板印象的激活则会得到一定的抑制，那么刻板印象的表达就会因此而被抑制；相反，低冲突背景下，对刻板印象的冲突监测不常发生，执行控制的水平会相应下降，这时就会很难抑制刻板印象的激活，则刻板印象表达就更易出现。有鉴于此，我们在采纳您建议的基础上，将文章的**题目修改为“基于不同冲突水平的认知控制对性别刻板印象表达的影响”**。

新增参考文献：

- Huntsinger, J. R., Sinclair, S., & Clore, G. L. (2009). Affective regulation of implicitly measured stereotypes and attitudes: Automatic and controlled processes. *Journal of Experimental Social Psychology, 45*(3), 560–566.
- Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129*(4), 522–544.

Krieglmeyer, R., & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *Journal of Personality and Social Psychology*, 103(2), 205–224.

**意见 2:** 问题提出中的术语使用比较随意、混乱，如刻板印象与偏见混用。

**回应:** 非常感谢审稿专家对核心概念的充分重视，对您的质疑我们做出如下回应。

**偏见与刻板印象既有联系又有区别。** 偏见是对社会群体及成员的正面或负面评价，而刻板印象则是与这些群体相关的知识、信念和期望的认知表征(Hamilton & Sherman, 1994; Amodio & Devine, 2006)。尽管学术界关于偏见和刻板印象的研究总是分开进行，但偏见和刻板印象是相互关联的(Mackie & Smith, 1998)。社会心理学的研究发现，刻板印象促进偏见，偏见的减少取决于刻板印象的改变。从这一角度来看，刻板印象是前因，偏见是后果。然而，有研究也发现偏见会影响刻板印象。如，高偏见的人通常比低偏见的人做出更多的刻板印象判断。尽管行为层面的研究表明，刻板印象和偏见是相互关联的。但神经影像学的证据却表明，在语义和情感记忆中，它们有不同的认知加工机制(Amodio & Devine, 2006; Amodio & Ratner, 2011)。在本研究中，我们研究的是性别刻板印象而非性别偏见。对于原文表述不到的地方我们已做了进一步完善。

**意见 3:** 反应抑制和主动抑制的突兀出现等等。话题也比较随意的切换。不清楚到底要研究什么，是主动控制对性别刻板印象的影响还是？一会在说认知控制，一会又在说主动控制。在问题提出甚至是全文中，到底在阐述和研究主动控制还是认知控制？如果是主动控制，有何理由研究？因为没有对比主动控制和反应控制的差异，更谈不上认知控制这个更大的概念。

**回应:** 非常感谢审稿专家细致入微的工作态度，但对您提出的该条建议，我们需要做出有必要的澄清。文章不存在概念随意混用的情况，但只要有读者质疑的声音，我们就有必要对文章相应内容的表述明晰化做进一步完善，感谢审稿专家为我们提供一个对于关键变量进行充分解释说明的机会。

**认知控制**指的是人们能根据目标或任务做出相应的反应，遇到分心物能保存目标、抑制习惯性反应或冲动行为的能力(Miller & Cohen, 2001)。**双重控制机制(DMC)理论**认为个体的认知控制能力具有一定的可变性，DMC的核心假设是，**认知控制**主要包括主动控制(proactive control)和反应控制(reactive control)两种模式(Braver et al., 2007)。**反应控制**强调了个体在单个认知事件中对目标刺激的感知判断(Brown et al., 2007)，能够对快速变化的外部条件作出

即时行动，通常代表一种短暂的局部或微观控制，也被称为晚期校正机制(Purmann et al., 2011)；**主动控制**则侧重于个体在整个认知事件中持续保持对目标信息的心理表征，以一种目标驱动的方式感知信息，对快速变化的外部条件做出更为持久的战略优化行动(Funes et al., 2010; Braver et al., 2009)，通常反应了一个较长的宏观时间尺度的发展模式(Braver et al., 2007; Miller & Cohen, 2001)，也被称为“**聚焦模式**”(Gratton et al., 1992)。从**起效的阶段和效果来看**，反应性控制依赖于干扰发生后的检测和解决(Braver, 2012)，在偏见出现后调整其认知模式，而主动控制依赖于干扰发生前的预期和预防，在偏见出现之前根据目标来确定其认知模式(Braver et al., 2009)。传统上，社会认知中偏见控制的理论模型强调的是一种**反应性的控制形式**，即当检测到偏见后所实施的认知加工方式，其作用是纠正激活的偏见对目标反应的影响(Fazio, 1990; Amodio et al., 2004; Sherman et al., 2008)。不同于反应控制，**主动控制**可以在激活人们头脑中关于刻板印象的相关信息之前，有效地将刻板印象的关注点迁移至任务目标本身，通过对任务困难的体验或对一个人的预期行为增加的动机来限制内隐刻板印象在行为中表达的可能性(Amodio et al., 2018)，而不仅仅在刻板印象出现或激活之后才开始发挥作用。

因此，本研究是基于**双重控制(DMC)理论**，即认知控制主要包括**主动控制(proactive control)**和**反应控制(reactive control)**两种模式(Braver et al., 2007)，进一步探讨不同冲突水平（高冲突、低冲突、基线三种条件）下诱发的认知控制（主动控制或反应控制），对刻板印象表达的影响。

新增参考文献：

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science, 15*(2), 88–93.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106–113.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanism of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory*. New York: Oxford University Press. 76–106

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 106*(18), 7351–7356.

- Brown, J. W., Reynolds, J. R., & Braver, T. S. (2007). A computational model of fractionated conflict-control mechanisms in task-switching. *Cognitive Psychology*, 55(1), 37–85.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The model as an integrative framework. *Advances in Experimental Social Psychology*, 23, 75–109.
- Funes, M. J., Lupiáñez, J., & Humphreys, G. (2010a). Analyzing the generality of conflict adaptation effects. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 147–161.
- Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121(4), 480–506.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24 (1), 167–202.
- Purmann, S., Badde, S., Luna-Rodriguez, A., & Wendt, M.. (2011). Adaptation to frequent conflict in the eriksen flanker task. *Journal of Psychophysiology*, 25(2), 50–59.
- Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, 115(2), 314–335.

**意见 4:** 是在研究刻板印象激活还是在研究刻板印象表达？

**回应:** 非常感谢审稿专家的执着追问。对您的疑问我们做出如下回应。

在研究中，我们采用的是**分类-确认范式而非启动-靶子一致性判断任务**。在分类任务中，被试只需要判断靶刺激属于哪一类别，在本研究中，要求被试判断图片是家务图片还是工作图片。我们考察的因变量是对靶子刺激分类的正确率和反应时，它能反映刻板化认知的行为结果而不能反映刻板化加工过程中某个瞬间的认知表征（如，刻板印象激活）。因此，我们研究的是刻板印象表达而非刻板印象激活。

**意见 5:** 研究者所罗列的假设是无法从问题提出中得来的。也就是说，研究者并没有有效地提出研究怎么样的关键性科学问题。

**回应:** 非常感谢审稿专家对于文字表述提出的期望与要求。基于这一要求，我们对引言的表述进行了进一步的完善。本次修改补充的内容均用**红色字体**对原文相应内容进行标识。

性别刻板印象是人们对两性认知的一种固化加工模式(Rudman & Kilianski, 2000)，它通常给人们提供了一种社会可接受性的标准，这种标准可以传达什么样的行为是适合男人的和

什么样的行为是适合女人的信息(Croft, 2016)。中国自古就有“男主外，女主内”的说法。国外也有研究表明，人们普遍具有男人应该养家糊口和女人应该照料家庭的性别刻板印象，这种刻板印象导致了对性别角色的不同期望及对性别发展技能上的不同侧重，进而促进了男女在社会行为上的分化差异(Eagly & Steffen, 1984; Eagly, 1987)。

传统的性别刻板印象对现代女性职业和家庭之间的协调有着很大的困扰，而男性也承担着来自家庭经济主力的巨大压力。这种刻板化的家庭角色分工对性别平等和家庭和谐往往会产生消极的影响(Cundiff & Vescio, 2016)。虽然在社会宏观层面，通过设置陪产假、育婴假、利用社会规范等干预措施可以来缓解男性和女性在职业和家庭中的角色分工并进而缓解公众所持有的性别刻板印象。但即便如此，深植于个体认知层面的性别刻板印象依旧没有被消除，并且有可能随时随地被激活。要避免刻板印象的自动激活是较为艰难和复杂的任务，而被自动激活的刻板印象却可以通过调整个体认知控制的方式对其表达进行有效抑制(Amodio & Swencionis, 2018)。

认知控制指的是人们能根据目标或任务做出相应的反应，遇到分心物能保存目标、抑制习惯性反应或冲动行为的能力(Miller & Cohen, 2001)。双重控制机制(DMC)理论认为个体的认知控制能力具有一定的可变性，这一理论提出认知控制主要包括主动控制(proactive control)和反应控制(reactive control)两种模式(Braver et al., 2007)。反应控制强调了个体在单个认知事件中对目标刺激的感知判断(Brown et al., 2007)，能够对快速变化的外部条件作出即时行动，通常代表一种短暂的局部或微观控制，也被称为晚期校正机制(Purmann et al., 2011)；主动控制则侧重于个体在整个认知事件中持续保持对目标信息的关注，一种目标驱动的感知信息的方式，对快速变化的外部条件做出更为持久的战略优化行动((Funes et al., 2010; Braver et al., 2009)，通常反应了一个较长的宏观时间尺度的发展模式(Braver et al., 2007; Miller & Cohen, 2001)，也被称为“聚焦模式”(Gratton et al., 1992)。例如，一个人在下班后总是在十字路口左转，但听说那条路上堵车，他可能在到达十字路口之前提醒自己右转，这便是主动控制策略；但他也可能采用反应控制模式，即到达十字路口后，遇见堵车，然后右转以避开拥挤的交通。从起效的阶段和效果来看，反应性控制依赖于干扰发生后的问题检测和解决(Braver, 2012)，它是在偏见出现后调整其认知模式，而主动控制依赖于干扰发生前的预期和预防，即在偏见出现之前根据目标来确定其认知模式(Braver et al., 2009)。

传统上，社会认知中偏见控制的理论模型强调的是一种反应性的控制形式，即当检测到偏见后所实施的认知加工方式，其作用是纠正激活的偏见对目标反应的影响(Fazio, 1990; Amodio et al., 2004; Sherman et al., 2008)。不同于反应控制，主动控制可以在激活人们头脑



中关于刻板印象的相关信息之前，有效地将刻板印象的关注点迁移至任务目标本身，通过对任务困难的体验或对一个人的预期行为增加的动机限制内隐刻板印象在行为中表达的可能性(Amodio et al., 2018)，而不仅仅在刻板印象出现或激活之后才开始发挥作用。几乎所有考察刻板印象的研究均离不开对比刻板印象一致信息和不一致信息的加工差异(贾磊等, 2016; Lai et al., 2016)，在实验室条件下常采用冲突任务来研究认知控制(Gonthier & Ambrosi, 2021)。冲突监控理论(conflict monitoring theory; van Veen & Carter, 2005)是最常见的对冲突任务进行理解的理论解释。该理论认为，对冲突信息进行加工包含着两个不同的成份：①评价成份(即冲突监测)，②执行成份(即执行控制)。不一致信息在认知过程中所产生的冲突强度是调节认知加工的基础，用于唤起与任务相关的知觉负载。其中，评价成分是对输入信息进行冲突识别并做出是否需要快速反应的决策成分，而执行成分是基于一定的任务指令，在知觉负载有限的条件下实现决策的成分(崔诣晨, 2016)。根据双重控制理论，人们在不同情境下会灵活地使用反应性控制或主动控制，从而影响刻板印象表达以及强度。在低冲突背景下，任务相对简单，对刻板印象的冲突监测不常发生，执行控制的水平会相应下降，难以抑制刻板印象的表达；而在高冲突背景下，任务困难，冲突监测会持续发生，执行控制的水平也适当提高，从而增强对任务目标本身的关注，促进主动控制模式的发生，进一步减弱刻板印象的表达(Amodio & Swencionis, 2018)。

Bartholow 和 Dickter (2008, 2010)的研究对种族分类任务中高冲突和低冲突试验的比率进行了操纵，其结果显示，高冲突组相对于基线组产生了一个减弱的刻板印象冲突模式，而低冲突组相对于基线组产生了一个更强的刻板印象冲突模式(Bartholow & Dickter, 2008; Dickter & Bartholow, 2010)，但这些研究并未直接验证认知控制对种族刻板印象表达的影响。在 Amodio 和 Swencionis(2018)的研究中，首次提出的主动控制模型(Proactive Model of Control) (图 1)，分别采用了武器识别范式(Payne, 2001)、刻板印象启动任务考察了主动控制对种族刻板印象表达的抑制作用，证实了相比基线组，个体在高冲突水平下会采用主动控制的加工模式，抑制种族刻板印象的表达，在低冲突水平下则会明显地表现出种族刻板印象。

前人的研究是在不同的任务情境中独立探究了反应性控制或主动控制对刻板印象表达的影响。并且，研究中所聚焦的种族刻板印象是西方社会所特有的一种文化现象，因而无法对双重控制机制理论提供充分的证据。有鉴于此，研究着眼于更具文化普适性的性别刻板印象，在同一实验任务中来探究认知控制影响刻板印象表达的机制，即探明反应性控制和主动控制在刻板印象表达中发挥作用的条件。

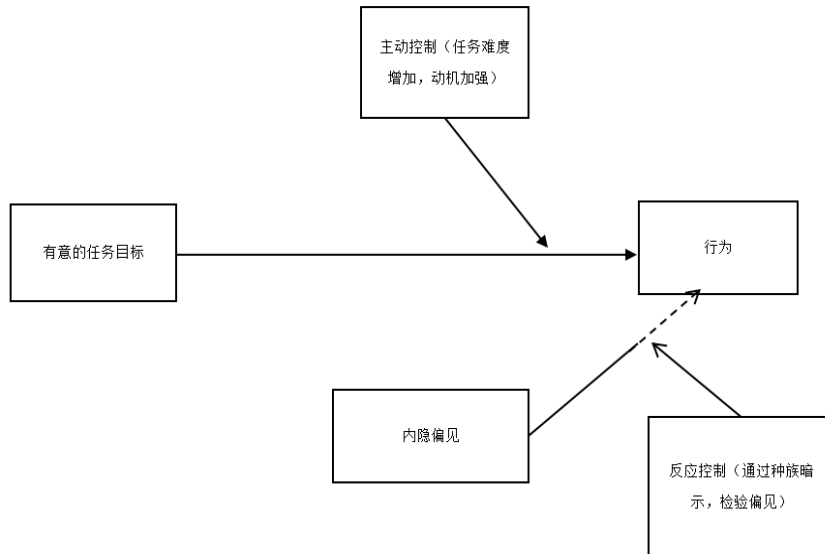


图1 主动控制模型 (Amodio & Swencionis, 2018)

性别刻板印象的表达与个体接触到的性别信息呈现形式相关。具体而言，性别信息呈现形式主要包括图片和语词两种类型。王沛等人(2010)提出，以具体的图片信息作为刺激，不论所选取的图片与某一群体的刻板联系有多紧密(如，黑人面孔-犯罪图片，女性面孔-家务图片等)，都只能反映出刻板印象的动态图式特点，却无法完全体现出刻板印象语义图式结构所暗含的抽象化加工特点；同样地，以抽象语义属性的文字信息作为目标刺激，则无法完全体现出刻板印象典型样例的具体化表征特点。情境-特异性启动效应(context-specific prime effects, CSPE)的研究表明，不同冲突比例的任务情境可以使个体对冲突信息或非冲突信息产生不同的关注程度，并根据冲突的特异性来持续启动不同的信息加工模式(Bailey et al., 2010)，在这一研究中，由于设定不同的冲突比例情境，因此可以更好地使知觉者调节对刻板化信息进行概念编码或知觉编码(袁菲, 2015; 崔诣晨, 2016)。此外，刻板印象的语义联结表现为自上而下的概念驱动加工，而具体样例联结则表现为自下而上的知觉驱动加工(Blair & Banaji, 1996)。因此，研究进一步探讨不同冲突水平引发的认知控制对性别刻板印象表达的影响是否具有跨信息呈现形式的一致性？

在现实生活中除了使认知控制策略在一定的条件下发挥作用以外，更重要的是在当前情况下引发的认知控制可以为后续的情况创造控制准备状态(Kleiman et al., 2014)，使得认知控制可以在不同的任务中进行及时调整。以往的研究普遍认为认知控制调整具有领域特异性，不同领域之间的控制调整是相当有限的(Egner, 2008; Hazeltine et al., 2011)。不过也有观点认为，认知控制引发的冲突调整状态具有领域适用性的特点(Kleiman et al., 2014; 袁菲, 2015)。对刻板印象的研究表明，前面试次中的经典 Flanker 任务的控制准备状态对后续刻板印象

Flanker 任务的表达起到了控制调整作用(Kleiman et al., 2014; 袁菲, 2015), 尽管已有研究试图说明认知控制对刻板印象表达的作用具有跨任务情境的一致性, 但由于前后实验试次中采用同一任务范式, 所以对于这一问题并未进行有效的解答。

综上所述, 本研究探究认知控制对性别刻板印象表达的作用机制。针对上述问题, 研究假设: (1)高冲突水平的目标分类任务会引发主动控制, 从而抑制性别刻板印象的表达; 低冲突水平的目标分类任务会引发反应性控制, 从而促进性别刻板印象的表达。(2)不同冲突水平引发的认知控制对性别刻板印象表达的抑制作用具有跨任务情境的一致性。

#### 参考文献:

- Bailey, K., West, R., & Anderson, C. (2010). A negative association between video game experience and proactive cognitive control. *Psychophysiology*, 47(1), 34–42.
- Bartholow, B. D., & Dickter, C. L. (2008). A response conflict account of the effects of stereotypes on racial categorization. *Social Cognition*, 26(3), 314–332.
- Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, 70(6), 1142–1163.
- Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113.
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive.
- Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), 7351–7356.
- Cundiff, J. L., & Vescio, T. K. (2016). Gender stereotypes influence how people explain gender disparities in the workplace. *Sex Roles*, 75(3–4), 1–13.
- Dickter, C. L., & Bartholow, B. D. (2010). Ingroup categorization and response conflict: interactive effects of target race, flanker compatibility, and infrequency on n2 amplitude. *Psychophysiology*, 47(3), 596–601.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, 23, 75–109.
- Funes, M. J., Lupianez, J., & Humphreys, G. (2010b). Sustained vs. transient cognitive control: Evidence of a behavioral dissociation. *Cognition*, 114(3), 338–347.
- Gonthier, C., Solène Ambrosi, & Agnès Blaye. (2021). Learning-based before intentional cognitive control:

- developmental evidence for a dissociation between implicit and explicit control. *Journal of Experimental Psychology Learning Memory and Cognition*.
- Hamilton, D. L., & Trolie, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In S. L. Gaertner & J. F. Dovidio (Eds.), *Prejudice, Discrimination, and Racism*. New York: Academic Press, 127–157.
- Jia, L., Zhu, S. R., Zhang, C. J., & Zhang, Q. L. (2016). The distributed semantic representation and activation processes of the implicit and explicit stereotypes: An examination based on the perspective of cognitive neuroscience. *Advances in Psychological Science*, 24(10), 1519–1533.
- Kleiman, T., Hassin, R. R., & Trope, Y. (2014). The control-freak mind: Stereotypical biases are eliminated following conflict-activated cognitive control. *Journal of Experimental Psychology: General*, 143(2), 498–503.
- Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., ... & Simon, S. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, 145 (8), 1001–1016.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24 (1), 167–202.
- Purmann, S., Badde, S., Luna-Rodriguez, A., & Wendt, M. (2011). Adaptation to frequent conflict in the eriksen flanker task. *Journal of Psychophysiology*, 25(2), 50–59.
- Yuan, F. (2016). *Cognitive monitoring in different contexts: The effect of conflict adaptation on gender stereotype*. (master dissertation, Shaanxi Normal University).

**意见 6:** 在具体的实验设计中自变量操纵、潜在变量控制、统计方法等诸多环节存在很大的纰漏甚至错误，缺乏清晰的统计分析的逻辑思路。例如：请报告 G-power 对于被试人数的预估计情况以及选用具体人数的理由。

**回应:** 感谢审稿专家的辛勤工作，我们对您的质疑做出如下回应和正文内容补充。由于文章在自检报告中报告了 G-power 对被试人数的预估计情况以及选用的具体人数，所以鉴于正文篇幅所限就未在原文中加以赘述。现根据您的建议，我们用红色字体标记部分对原文进行了补充，内容如下：

### 2.1.1 被试

使用 G-power 3.1 (Faul et al., 2009) 计算研究所需的样本量，效应量设置为 0.25，设定  $\alpha$

为 0.05, 并且检验效能为 0.8, 至少需要 36 人。基于先前同类研究中所用到的样本量为 25 ~ 63 之间(Amodio & Swencionis, 2018; Kleiman et al., 2016), 以及综合考虑到实验过程中可能出现的问题(如无效被试、男女人数差异等), 实际招募被试 60 名(男性 26 名), 平均年龄 24.60 岁(SE=2.82), 裸视或矫正视力正常, 能熟练操作计算机, 之前未参加过类似实验。所有被试均为自愿参加, 实验结束后给予一定的报酬。

参考文献:

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

### 3.1.1 被试

**同实验 1, 最低被试量需要 36 名。**实际招募 62 名被试参与实验, 其中两名被试未能遵循任务指示被排除在外, 三名被试的数据因无效反应过多被剔除, 两名被试因反应的正确率过低被排除, 最终有效被试共 55 名(男性 26 名), 平均年龄 19.87 岁(SE=2.54)。所有被试视力或矫正视力正常, 能熟练操作计算机, 之前未参加过类似的实验。所有被试均为自愿参加, 实验结束后可获得相应的报酬。

### 4.1.1 被试

**同实验 1 和实验 2, 最低被试量需要 36 名。**实际招募 67 名被试参与实验, 其中 4 名被试未能完成全部实验流程被排除在外, 5 名被试的数据因无效反应过多被剔除。最终有效被试共 58 名(男性 29 名), 平均年龄 22.33 岁(SE=2.67)。所有被试视力或矫正视力正常, 能熟练操作计算机, 之前未参加过类似的实验。所有被试均为自愿参加, 实验结束后可获得相应的报酬。

**意见 7:** 实验 3 同实验 1 和实验 2 之间有何必然的联系? 在前面的实验的基础上要解决什么问题? 因为实验 3 的范式和刺激结构完全同前面的实验不同。

**回应:** 非常感谢审稿专家的辛勤工作。实验 3 同实验 1 和 2 之间的逻辑如下。

实验 1 和实验 2 从横向的不同任务情境(图片分类任务和语词分类任务)的设置中探究不同冲突水平引发的认知控制对性别刻板印象表达的作用是否具有跨任务情境的一致性。实验 3 是从纵向不同任务情境的设置中, 利用改进的试次-试次控制调整范式(trial to trial control adjustment paradigms; Kleiman et al., 2014), 每个完整试次都由两个连续的试次组成, 共包含两个不同的行为反应, 前一个试次是使用实验 2 中使用到的词语分类任务, 接着第二个任务是性别 Flanker 任务, 来考察不同冲突水平引发的认知控制对性别刻板印象表达的作用

否具有跨任务情境的一致性。

在现实生活中除了使认知控制策略在一定的条件下发挥作用以外,更重要的是在当前情况下引发的认知控制可以为后续的情况创造控制准备状态(Kleiman et al., 2014),使得认知控制可以在不同的任务中进行及时调整。以往研究普遍认为认知控制调整具有领域特异性,不同领域之间的控制调整是相当有限的(Egner, 2008; Hazeltine et al., 2011)。不过也有观点认为,认知控制引发的冲突调整状态具有领域适用性的特点(Kleiman et al., 2014; 袁菲, 2015)。有关刻板印象的研究表明,之前试次中的经典 Flanker 任务引起的冲突适应可以对后续 Flanker 任务中刻板印象的表达起到跨领域的控制调整作用(Kleiman et al., 2014; 袁菲, 2015)。

因此,实验 1 和实验 2 横向考察不同任务情境中冲突引发的认知控制对性别刻板印象表达的作用是否具有跨任务情境的一致性。实验 3 纵向考察不同任务情境中冲突引发的认知控制对性别刻板印象表达的作用是否具有跨任务情境的一致性。这样的设计能更全面系统地揭示认知控制对刻板印象表达的作用机制。

意见 8: 实验 1-3 做的是刻板印象激活效应而非刻板印象的表达,因为不存在表达的任务。这一点恰恰与研究者在问题提出与讨论等部分提到的自己研究主题不符。

回应: 非常感谢审稿专家细致入微的工作态度,但对您提出的该条建议,我们需要做出有必要的澄清。

在本研究中,实验 1-3 采用的是分类-确认范式而非启动-靶子一致性判断任务。在分类任务中,被试只需要判断靶刺激属于哪一类别,比如本研究中是判断家庭还是工作图片。因此,刻板印象激活虽是前提但不能被行为直接观测,但刻板印象表达却是可被观测的行为结果。冲突检测理论提出,对冲突信息进行加工包含着两个不同的成份:①评价成份(即冲突监测),②执行成份(即执行控制)。不一致信息在认知过程中所产生的冲突强度是调节认知加工的基础,用于唤起与任务相关的知觉负载。正如在本研究中,在高冲突背景下,冲突监测会持续发生,执行控制的水平也会适当提高,这时刻板印象的激活则会得到一定的抑制,那么刻板印象的表达就会得到较强的抑制;相反,低冲突背景下,对刻板印象的冲突监测不常发生,执行控制的水平会相应下降,这时就会很难抑制刻板印象的激活,则刻板印象表达就更易出现。此外,我们考察的因变量是对目标刺激分类的正确率和反应时,而不是启动和目标的一致性判断过程中刻板印象激活瞬间的认知表征。因此,研究中被观测的是刻板印象的表达而非激活。

意见 9：实验 1 中启动刺激仅各自用了 2 张男性-女性照片，这里练习（熟悉）效应对于实验结果的影响是巨大的，如何排除？选取了个 6 张照片，与启动刺激进行配对后在每种刺激条件下仅有 12 张，加上不断重复，这样小而熟悉的刺激集合所得到的结果如何确保统计可靠性？

回应：非常感谢审稿专家对文章的细致分析，对您的疑惑我们具体回复如下。

在 Amodio 等(2018)的研究中，采用武器识别范式，以两张白人面孔和两张黑人面孔作为启动图片，四把手枪和四个手动工具为目标刺激，探究主动控制对内隐种族偏见的影响。同时，在 Chen 等人(2021)的研究中，也是选取两张男性面孔图片和两张女性面孔图片作为启动刺激，目标刺激包括四张注射器图片和四张笔的图片，探讨在不同情绪状态下，冲突驱动的主动控制加工对内隐职业性别刻板印象表达的影响。另外，研究者考虑到实验的任务是被试判断目标刺激与前面呈现的面孔图片是否一致。即使男女面孔图片各只有 2 张，被试非常熟悉面孔图片，但被试也要根据目标刺激判断与之前的启动刺激图片是否一致。因此，在本研究中，实验 1 沿用 2 张男性面孔和 2 张女性面孔作为启动刺激，4 张家务图片和 4 张工作图片。但是研究 1 出现阴性结果，研究者考虑到可能存在熟悉效应，所以在 Amodio 等(2018)的研究基础上，实验 2 和实验 3 启动刺激采用四张男性面孔和四张女性面孔图片。

参考文献：

Chen, L., Zeng, S, Y., & He, R. (2021). Negative Emotions Can Interfere with the Inhibitory Effect of Proactive Control on Implicit Stereotypes. *North American Academic Research Journal*, 4(4),53–69.

意见 10：在实验 1 中如何控制个体本身的初始情绪与性别角色的观念水平这两个研究者人为地潜在变量？为什么没有数据统计的报告？仅仅把性别角色的观念水平作为协变量进行控制是否足够？在后面的两个实验中为什么对这些控制变量的控制变得更弱？

回应：非常感谢审稿专家的细致阅读。针对审稿专家提出的问题，对此问题做出如下回应：

首先，本研究中实验 1、2、3 测量情绪仅仅是为了剔除有极端情绪的被试；其次，有研究表明性别角色观念会影响性别刻板印象的表达，因此本研究将性别角色观念作为协变量进行控制，因数据统计结果未在原文中报告，现已添加至原文中，用红色字体相应标出。最后，实验 1、2、3 均在实验前测量了被试的情绪和性别角色观念水平，鉴于正文篇幅所限未在原文中加以赘述，只在文章实验 1 中报告，现已添加至原文中，因此，不存在控制变量更弱的情况。

#### 2.1.4 实验程序

由于个体的性别角色的观念水平会对刻板印象的认知加工产生影响。结合中国妇女社会



地位调查(CWSSS)和中国综合社会调查(CGSS)中的性别观念的题项形成《性别角色观念量表》，将其区分为传统主义取向和现代主义取向(Rajadhyaksha et al., 2015)。该实验中，《性别角色观念量表》的信度系数为 0.83，结构效度为 0.67，60 名被试在该量表中的得分为  $M=3.80$ ， $SD=0.51$ ，并且在测量家庭角色分工的题项中，父亲主外母亲主内的比例为 53.3%，母亲主外父亲主内为 3.3%，父母双方处于均衡状态的人数为 43.3%。

## 2.2 结果

对每种冲突条件下刺激类型的平均正确率进行 3(冲突水平：高冲突，基线，低冲突)×2(启动刺激：男性图片，女性图片)×2(目标刺激：工作图片，家庭图片)的重复测量方差分析，结果显示：启动刺激和目标刺激之间的交互作用显著， $F(1, 59)=4.91$ ， $p=0.031$ ， $\eta_p^2=0.08$ 。其他的主效应和交互作用均不显著( $ps>0.05$ )。

对每种冲突条件下刺激类型的反应时均值进行 3(冲突水平：高冲突水平，基线水平，低冲突水平)×2(启动刺激：男性图片，女性图片)×2(目标刺激：工作图片，家庭图片)的重复测量方差分析，结果并未显示出任何显著的效应( $ps>0.05$ )。

将性别角色观念均分作为协变量，在冲突水平上对被试反应的平均正确率和反应时分别进行重复测量方差分析，结果冲突水平的主效应也不显著( $ps>0.05$ )。

### 3.1.4 实验程序

测量被试的情绪状态和性别角色观念水平的问卷同实验 1，该实验中，《性别角色观念量表》的信度系数为 0.79，结构效度为 0.62，55 名被试在该量表中的得分为  $M=3.89$ ， $SD=0.49$ ，并且在测量家庭角色分工的题项中，父亲主外母亲主内的比例为 47.3%，母亲主外父亲主内为 0，父母双方处于均衡状态的人数为 49.1%，父母均不兼顾的人数为 3.6%。实验整体流程与实验 1 基本一致，不同之处是实验 2 的任务是对目标词语做出属于“家庭”或“职业”的分类判断，具体的试次的流程如图 3 所示。

## 3.2 结果

将性别角色观念均分作为协变量，在冲突水平上对被试反应的平均正确率和反应时分别进行重复测量方差分析，正确率的结果显示，冲突水平的主效应依然显著， $F(2, 106)=4.99$ ， $p=0.014$ ， $\eta_p^2=0.09$ ；反应时的结果同样显示，冲突水平的主效应也依然显著， $F(2, 106)=3.31$ ， $p=0.04$ ， $\eta_p^2=0.06$ 。

### 4.1.2 实验设计

测量被试情绪状态和性别角色观念水平的问卷同实验 1，该实验中，《性别角色观念量表》的内部一致性信度为 0.84，结构效度为 0.71，58 名被试在该量表中的得分为  $M=3.76$ ，



$SD=0.50$ ，并且在测量家庭角色分工的题项中，父亲主外母亲主内的比例为 51.7%，母亲主外父亲主内为 5.2%，父母双方处于均衡状态的人数为 41.4%，父母均不兼顾的人数为 1.7%。

## 4.2 结果

将性别角色观念均分作为协变量，在冲突水平上对被试反应的平均正确率和反应时分别进行重复测量方差分析，正确率的结果显示，冲突水平的主效应依然显著， $F(2, 112)=9.08$ ， $p=0.001$ ， $\eta_p^2=0.14$ ；反应时的结果同样显示，冲突水平的主效应也依然显著， $F(2, 112)=5.78$ ， $p=0.004$ ， $\eta_p^2=0.09$ 。

**意见 11：**实验 1 中练习正确率达到 70%即为有效数据有何依据，为什么不设置到 90%甚至更高水平？有何依据？

**回应：**非常感谢审稿专家的仔细审阅，对您的疑惑我们做出如下回应。

练习实验主要是让被试熟悉任务以及相应的按键反应，本研究中设置练习 70%的实验正确率，目的是保证被试明白实验意图，但如果练习正确率设置 90%或者更高的正确率，则可间接说明被试的判断能力较为同质，难以体现因样本差异对冲突水平做出不同的反应所带来的影响。

**意见 12：**实验 1 之所以缺乏显著的结果，一个重要的潜在污染可能是图表的高度可以识别性。

**回应：**感谢审稿专家提出的意见，按照您的意见，我们做出如下说明和回应：

实验 1 出现阴性结果，这一结果可能的原因包含**实验材料的外在表现维度、实验任务的性质以及内在加工形式**这几个层面。

第一，实验 1 是**图片分类任务**，研究者考虑到高清图片具有很高的生态效度，所以启动刺激和目标刺激均为彩色高清图片，但这种形式的图片也相应的具有易辨识、易记忆、易加工的特性(Yuan et al., 2019)。其次，启动图片和目标图片较少，重复配对试次较多，使得被试在习得材料的加工模式后很难受不同比例试次的冲突水平的的影响。另外，大量有关启动刺激的研究发现刺激材料的丰富性易弱化练习效应的作用，加强实验处理对结果的影响(Kidder et al., 2018; Schmidt, 2019)。

第二，实验任务方面。与测量刻板印象的经典任务——启动-靶子一致性判断任务不同，图片分类任务中强调被试需要判断靶刺激的类别(如，家务图片或工作图片)，而启动-靶子一致性判断任务则需要对启动刺激和靶子刺激(目标刺激)两者的配对做一致性和不一致性

的匹配判断。从难度和经验上来说，显然完成前一个任务所需的难度更小，目标刺激受启动刺激的影响也更小，所以即使可以体验到不同的冲突变化，也会容易出现冲突适应效应，使得不同冲突水平的结果没有显著差异。

最后，关于文字和图像处理理论观点认为：图像更倾向于自下而上的具体样例表征加工，而文字加工则包含了比图片加工更多的抽象表征加工(Yuan et al., 2019)，图像在认知层面减少了自上而下的转换资源，可直接依据具体情境来做出判断(Sherman, 1996)，因此可能会出现处理条件对实验结果没有影响，即图片在不同冲突水平下的加工速度和准确性并无优劣之分。

参考文献：

Yuan, J., Tian, Y., Huang, X., Fan, H., & Wei, X. (2019). Emotional bias varies with stimulus type, arousal and task setting: meta-analytic evidences. *Neuroscience & Biobehavioral Reviews*, 107, 461–472.

Kidder, C. K., White, K. R., Hinojos, M. R., Sandoval, M., & Crites, S. L. (2018). Sequential stereotype priming: A meta-analysis. *Personality and Social Psychology Review*, 22(3), 199–227.

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: an updated review. *Psychonomic Bulletin & Review*, 26, 753–771.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality & Social Psychology*, 70(6), 1126–1141.

**意见 13：**不同水平下的主动控制与反应控制估算值是如何得出的，这样的数据就可以说明不同认知控制的水平了吗？

**回应：**感谢审稿专家提出的意见。根据审稿专家的宝贵意见，对此问题做出如下回应。

本研究中使用加工分离程序(Process Dissociation Procedure, PDP)验证我们的假设(Jacoby, 1991; Jacoby et al., 1999)。**PDP 是一种从行为模式中评估控制加工过程(即任务目标一致)和自动加工过程(即刻板印象偏见一致)对任务绩效的独特贡献的方法。**在 PDP 框架中，控制的估计(PDP-C)表示一个人将以一种准确的、目标一致的方式作出反应的可能性，而不存在性别启动引起的刻板印象驱动的刻板印象(P[一致试验的正确反应]-P[不一致试验的错误反应])。自动性估计(PDP-A)表示这样一种可能性：在控制失败程度上，一个人的反应会因为与目标的刻板印象联系而受到性别启动的刻板印象(P[刻板印象错误/(1-控制)])；(Payne, 2001)。**例如，对实验 2 的数据分析过程如下：**

首先，对 PDP-C 的估计值进行 3(冲突水平：高冲突水平，基线水平，低冲突水平)×2(启

动刺激：男性图片，女性图片)的重复测量方差分析，结果显示，冲突水平主效应显著， $F(2, 108)=7.94$ ， $p=0.001$ ， $\eta_p^2=0.13$ ，但启动刺激的主效应不显著，两者的交互作用也不显著( $ps>0.05$ )。也就是说，不同冲突水平的控制加工估算值不因男性和女性启动图片类型而异。进一步多重比较结果表明，高冲突水平下的控制加工估算值( $M=0.93$ ， $SD=0.01$ )显著高于低干扰水平( $M=0.88$ ， $SD=0.01$ )和基线水平( $M=0.90$ ， $SD=0.01$ )， $p<0.001$ ， $p=0.019$ 。不过基线水平与低干扰水平之间没有明显的差异( $p=0.228$ )。高冲突水平引发了最高程度的控制加工，低冲突水平下的控制加工最少，基线水平的控制加工估算值位于中间。这个分析与前一节中报告的冲突水平对准确率的主要影响基本相同，**PDP-C** 本质上表示的是冲突水平作为启动刺激函数的平均准确率。这种模式也进一步支持了我们的假设，高冲突的加工任务更容易引发主动控制的加工。

其次，我们分析了 **PDP-A** 的估算值，以验证我们的预测。由于 PDP-A 公式在 PDP-c=1 时无法求解(即，当一个或多个试次类型的准确率为 100%时)，会造成有些受试者的 PDP-a 的分值缺失。仅有 20 名被试在所有试次条件下都有有效的 PDP-A 分值，将这 20 名被试纳入本次分析中。对的估算值做 3(冲突水平：高冲突水平，基线水平，低冲突水平)×2(启动刺激：男性图片，女性图片)的重复测量方差分析，结果并未出现任何显著的主效应和交互作用( $ps>0.05$ )，不过数据显示，PDP-A 的估算值在高冲突条件下的自动估算值最高( $M=0.69$ ， $SD=0.05$ )，基线水平位于中间( $M=0.63$ ， $SD=0.05$ )，低冲突水平最低( $M=0.56$ ， $SD=0.06$ )，见表 4。这样的模式与冲突效应应该在不影响内隐联想的情况下调节控制加工的参与是一致的(Hilgard et al., 2015)。

#### 参考文献：

- Jacoby, L. L. (1991). a process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541.
- Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In S. Chaiken & Y. Trope (Eds.), *Dual Process Theories in Social Psychology*. New York, NY: Guilford Press.383–400
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81(2), 181–192.
- Hilgard, J., Bartholow, B. D., Dickter, C. L., & Blanton, H. (2015). Characterizing switching and congruency effects in the Implicit association Test as reactive and proactive cognitive control. *Social Cognitive and affective Neuroscience*, 10(3), 381–388.

意见 14: 为什么在实验 2 出现了各 4 张图片的启动刺激? 而实验 1 则是各 2 张? 按理说图片加工条件需要更多的启动图片以降低熟悉效应。

回应: 感谢审稿专家提出的意见。根据审稿专家的宝贵意见, 对此问题做出如下回应。

在 Amodio 等(2018)的研究中, 采用武器识别范式, 以两张白人面孔和两张黑人面孔作为启动图片, 四把手枪和四个手动工具为目标刺激, 探究主动控制对内隐种族偏见的影响。同时, 在 Chen 等人(2021)的研究中, 也是选取两张男性面孔图片和两张女性面孔图片作为启动刺激, 目标刺激包括四张注射器图片和四张笔的图片, 探讨在不同情绪状态下, 冲突驱动的主动控制加工对内隐职业性别刻板印象表达的影响。另外, 研究者考虑到实验的任务是被试判断目标刺激与前面呈现的面孔图片是否一致。即使男女面孔图片各只有 2 张, 被试非常熟悉面孔图片, 但被试也要根据目标刺激判断与之前的启动刺激图片是否一致。因此, 在本研究中, 实验 1 沿用 2 张男性面孔和 2 张女性面孔作为启动刺激, 4 张家务图片和 4 张工作图片。但是研究 1 的出现阴性结果, 研究者考虑到可能存在熟悉效应, 所以在 Amodio 等人(2018)的研究基础上, 实验 2 和实验 3 启动刺激采用四张男性面孔和四张女性面孔图片。但并不是启动刺激越多越好, 首先启动刺激越多, 会有其他的干扰效应, 比如面孔吸引力等; 其次, 本研究主要探讨不同冲突水平的认知控制对性别刻板印象表达的影响, 因此设计启动刺激的数量和目标刺激数量固定, 只是不同的匹配, 所以冲突水平是独立于男女面孔而被操纵(Amodio et al., 2018)。

意见 15: 实验 2 为什么仅仅试图控制性别角色观念水平, 而且与实验 1 的控制方法不同?

回应: 感谢审稿专家提出的意见。根据审稿专家的宝贵意见, 对此问题做出如下回应。

实验 1、2、3 均测量了被试的初始情绪和性别角色观念。但鉴于正文篇幅所限未在原文中加以赘述, 现将原文部分修改如下, 本次修改补充的内容均用红色字体对原文相应内容进行标识。

#### 3.1.4 实验程序

测量被试的情绪状态和性别角色观念水平的问卷同实验 1, 该实验中, 《性别角色观念量表》的信度系数为 0.79, 结构效度为 0.62, 55 名被试在该量表中的得分为  $M=3.89, SD=0.49$ , 并且在测量家庭角色分工的题项中, 父亲主外母亲主内的比例为 47.3%, 母亲主外父亲主内为 0, 父母双方处于均衡状态的人数为 49.1%, 父母均不兼顾的人数为 3.6%。实验整体流程与实验 1 基本一致, 不同之处是实验 2 的任务是对目标词语做出属于“家庭”或“职业”的

分类判断，具体的试次的流程如图 3 所示。

#### 4.1.2 实验设计

测量被试情绪状态和性别角色观念水平的问卷同实验 1，该实验中，《性别角色观念量表》的内部一致性信度为 0.84，结构效度为 0.71，58 名被试在该量表中的得分为  $M=3.76$ ， $SD=0.50$ ，并且在测量家庭角色分工的题项中，父亲主外母亲主内的比例为 51.7%，母亲主外父亲主内为 5.2%，父母双方处于均衡状态的人数为 41.4%，父母均不兼顾的人数为 1.7%。

意见 16：实验 2 在对冲突水平正确率的主效应做进一步的事后多重比较的结果存疑：0.94、0.95、0.96 到底有何差异一定会到显著性水平？那么“但低冲突水平下四种刺激类型的平均正确率( $M_{低冲突}=0.94$ ， $SE=0.01$ )与基线水平差异不显著( $M_{基线}=0.95$ ， $SE=0.01$ ， $p=0.228$ )”这个结果又做何解释？

回应：感谢审稿专家细致审阅。我们对您的疑惑做出如下回应：

首先，我们对原始数据重新检查并分析，仍然发现跟原文中一样的结果， $p=0.001$ ， $\eta_p^2=0.13$ ，且我们查阅文献发现效应量  $\eta_p^2$  大于 0.5，属于大效应，即我们的结果具有实际意义，可以反映实验的真实效果。其次，因为高冲突、基线水平以及低冲突的平均值 0.96、0.95 和 0.94 保留了两位小数，很难区分差异，因此在下文中我们保留三位小数。

在实验 2 中，冲突水平主效应依然显著， $F(2, 108)=7.94$ ， $p=0.001$ ， $\eta_p^2=0.13$ ；进一步的事后多重比较，如表 3 结果显示：高冲突水平下四种刺激类型的平均正确率( $M_{高冲突}=0.962$ ， $SD=0.01$ )高于低冲突水平( $M_{低冲突}=0.942$ ， $SD=0.01$ ， $p<0.001$ )和基线水平( $M_{基线}=0.948$ ， $SD=0.01$ ， $p=0.019$ )；但低冲突水平下四种刺激类型的平均正确率( $M_{低冲突}=0.942$ ， $SD=0.01$ )与基线水平差异不显著( $M_{基线}=0.948$ ， $SD=0.01$ ， $p=0.228$ )。因此，正如预测的那样，被试在高冲突水平上表现最好。

表 3 不同冲突水平下对目标刺激进行分类判别的正确率及反应时

	高冲突 $M(SD)$	基线 $M(SD)$	低冲突 $M(SD)$
正确率(%)	<b>0.962 (0.01)</b>	<b>0.948 (0.01)</b>	<b>0.942 (0.01)</b>
反应时(ms)	467.78 (9.77)	456.70 (8.13)	442.45 (10.17)

参考文献：

郑昊敏，温忠麟，吴艳. (2011). 心理学常用效应量的选用与分析. *心理科学进展*, 19(12):1868-1878.

意见 17: 实验 2 三阶交互作用的数据量极低, 进一步进行统计分析的结果没有什么实质性意义。

回应: 非常感谢审稿专家细致入微的工作态度, 但对您提出的该条建议, 我们需要做出有必要的澄清。

实验 2 中冲突水平、启动刺激和目标刺激三者之间交互作用显著,  $F(2, 108)=3.65$ ,  $p=0.029$ ,  $\eta_p^2=0.06$ 。虽然三阶交互作用  $p$  值并不是小于 0.01 或者 0.001, 但是  $p=0.029$ , 小于 **0.05**, 满足假设检验 I 类错误小于 0.05 的要求, 因此可以拒绝原假设, 接受备择假设。并且具有显著的统计学的差异。

意见 18: 讨论部分结果阐述与解释混杂在一起, 有的解释与结论完全与研究结果无关, 难以让人心信服。

回应: 感谢审稿专家提出的意见。根据审稿专家的宝贵意见, 我们对讨论部分进行了补充修改, 本次修改补充的内容均用**红色字体**对原文相应内容进行标识。

### 3.3 讨论

从词语分类任务正确率的结果来看, 高干扰引发的控制加工优势在任务准确性上得到了很好的检验, 即主动控制通过将注意力从与任务无关的干扰物(如性别)转移, 以促进对高冲突比例任务的整体目标的处理(Amodio & Swencionis, 2018)。在高冲突条件下, 当男性图片/女性图片作为启动刺激出现时, 被试对代表家庭和工作的词语均能做出正确的判断; 而在基线和低冲突条件下则出现了典型的性别刻板印象模式, 即当男性图片作为启动刺激出现时, 被试表现出了对代表工作的词语更为准确的判断, 而当女性图片作为启动刺激出现时, 表现出了对代表家务的词语更为准确的判断。此外, 高冲突水平和基线水平下的反应时高于低冲突水平条件下的反应时。利用加工分离程序对不同冲突水平下的控制加工进行分析发现, 高冲突和基线水平相应的控制加工估算值较高, 低冲突水平相应的控制加工估算值较低。因此, 我们推测, 低冲突水平下可能增加了参与者对反应性控制的依赖, 而高冲突水平下被试更依赖于主动控制。不过, 双重认知控制理论认为反应性控制和主动控制的认知加工过程可以根据情况发生动态的变化(Braver, 2012; Schmid et al., 2015)。但当前的研究发现, 高冲突和基线水平的自动加工估算值较高, 而低冲突水平相应的控制加工估算值较低, 间接证明了反应性控制和主动控制在同一任务中的协同加工的性质(Funes et al., 2010)。

在本研究中, 我们对家庭角色分工的调查结果得出, 大多数家庭中“父亲更倾向于挣钱养家, 母亲更倾向于照顾家庭”, 这一角色分工的差异性更符合低冲突任务中的性别刻板印

象的呈现方式。而在高冲突任务中，刻板印象一致性刺激出现的比例很低，冲突刺激出现的比例更高，这种反性别刻板印象的呈现方式在现实生活中并不常见。根据刻板印象的混合模型的观点(Sherman, 1996)，低冲突条件的文字刺激由于认知经济的原则会更倾向于抽象表征，相应的，高冲突条件和基线条件则由于配对刺激的特异性更符合样例表征，而样例表征的特点即是准确且加工速度较慢。

### 4.3 讨论

实验 3 的主要目的是为了证明不同冲突水平引发的认知控制对性别刻板印象表达的作用是否具有跨任务情境的一致性，因此确保任务转换中前一任务在不同冲突水平下的认知控制差异是验证这一假设的关键前提(Egner, 2008; Freitas et al., 2007)。采用试次-试次控制调整范式发现，在前一个试次的词语分类任务中，相较于基线水平，高冲突条件下引发主动控制，正确率更高，因此重复了实验 2 的结果。进一步分析不同冲突水平的分类任务对后续性别 Flanker 任务的加工模式，我们发现，高冲突水平分类任务中激活的主动控制加工可诱发对后续任务的认知控制调整，即在高冲突分类任务后，被试在性别 Flanker 任务的正确率更高；并且，通过加工分离程序计算出的高冲突水平的控制加工估算值比低冲突和基线水平平均高。

实验 2 中词语分类任务高冲突条件引发的主动控制的加工优势又一次得到了验证，也为跨任务情境的一致性提供了基础(Schmidt, 2019)，这为内隐刻板印象任务可能会改变冲突监测和认知控制的调整提供了有力的支撑(Amodio et al., 2008; Freitas et al., 2007)。

## 5 总讨论

不同冲突水平的任务背景和不同条件诱发的认知控制对性别刻板印象表达有着不一样的效果。实验 1 的结果表明，无论是何种冲突试次的比例水平，图片分类任务均能抑制性别刻板印象的表达；而对于实验 2 的词语分类任务而言，相较于基线水平，高冲突水平下性别刻板印象的表达受到了抑制，低冲突水平下性别刻板印象的表达得到了激活；实验 3 的结果则进一步表明高冲突水平下产生的主动控制能够在随后的任务中被持续保持，不受任务类型变化的影响。这些发现暗示更大的冲突试次比例可能会对控制刻板印象的表达更有效，而控制的有效性和刻板印象的具体呈现形式有关，个体可以敏感觉察出冲突大小的精细变化，且高冲突任务引发的冲突适应为后续任务提供了控制准备状态(Bukowski et al., 2019; Schmidt, 2019)。

### 5.1 任务冲突比例对性别刻板印象表达的影响



经典的无社会意义的认知冲突实验研究已表明,冲突试次占比高的实验条件可诱发一种主动控制模式(Gratton et al., 1992)。实验 2 和实验 3 使用加工分离程序的分析也发现,相对于低冲突条件,高冲突条件下的控制加工值(PDP-C)更大(基线时的控制估计为中间值)。这间接说明了高冲突水平诱发主动控制,而不是更多性别刻板印象的反应性激活。从常识上来看,更强的干扰环境导致判断准确性的提高这一发现可能有违直觉,但确与主动控制的机制以及过去的研究有着直接的相似性(Appelbaum et al., 2014)。通过进一步对自动加工值(PDP-A)的数据处理,我们发现,主动控制和反应控制可以协同作用于内隐刻板印象的行为表达,而不是刻板印象的来源(如自动激活的刻板印象)。另有研究者发现,主动控制与反应控制并非独立存在,在知觉到冲突有所变化时,个体可同时运用两种认知控制策略以形成正确的反应(Bugg & Braver, 2016),实验结果也充分论证了主动控制和反应控制能根据情况需求产生一定互补性的观点。

## 5.2 任务冲突比例下,刻板印象的抑制表达有其限制性

从实验 1 的分析结果上来看,不同试次比例的冲突任务并没有使个体的认知加工出现明显的差异,出现这种结果的原因可能与实验材料的外在表现维度及内在加工形式有关。实验 2 将图片更换为词语作为目标刺激的呈现形式,结果表明高冲突条件下的性别图片呈现时,抽象文字类型的目标刺激引发的主动控制对性别角色刻板印象的表达有一定的抑制作用。

首先,实验 1 中的启动刺激和目标刺激均为彩色高清图片,具有很高的生态效度,但这种形式的图片也相应的具有易辨识、易记忆、易加工的特性(Yuan et al., 2019),并且关于文字和图像处理区别的理论观点认为:图像更倾向于自下而上的具体样例形式加工,而文字加工则包含了比图片加工更多的抽象形式加工(Yuan et al., 2019),图像在认知层面减少了自上而下的转换资源,可直接依据具体情境来做出判断(Sherman, 1996),因此可能会出现不同冲突水平对实验结果没有影响的结果,即图片在不同冲突水平下的加工速度和准确性并无优劣之分。实验 2 验证了这样一个假设,即相对于弱干扰条件,强干扰的背景下会引起更多的主动控制,进而减少内隐刻板印象的表达(Amodio & Swencionis, 2018)。从实验 1 和实验 2 的结果来说,不同类型的目标刺激材料在高干扰环境下所引发的主动控制和反应性控制加工并不是完全一致的,进一步为不同冲突水平对信息呈现形式的敏感性存在差异提供了有效证据。

## 5.3 任务冲突比例下,刻板印象的抑制表达具有领域适用性的特点

实验 3 在实验 2 的结论基础上,利用改进的试次-试次控制调整范式证明先前高冲突比例的性别刻板印象分类任务与后续性别 Flanker 任务有着相似的认知控制模式。最近一些研

究也充分证明冲突状态能够为后续任务提供控制准备,进而引发认知控制的动态加工中更强的冲突适应(Kleiman et al., 2014; Bukowski et al., 2019; 张孟可等, 2021)。Bukowski 等人(2019)为了评估认知灵活性,使用包含社会类别的任务转换程序,在实验 1 中,被试按性别或年龄分类,在实验 2 和实验 3 中,被试按性别或社会角色分类,同样也发现被试的认知灵活性主要表现在引起更多认知冲突的任务情况下,并且认知灵活性在不同的任务冲突比例中体现为控制策略的灵活调整。因此,被试在高冲突背景下的性别 Flanker 任务中,更易引发主动性控制来抑制 Flanker 刺激诱发的干扰,将注意力集中在位于中央位置的目标刺激上。低冲突背景下则更依赖于反应性控制,从整个刺激中提取信息并做出反应。

综合而言,以往大量的研究关注了认知控制灵活性的转变,但是都局限在相同的冲突任务背景内进行比较,也常忽略信息表征的特殊表现形式和社会刺激的不同类型之间的控制准备状态。由于我们日常各类行为活动都是在不同比例的社会冲突环境中进行,因而考虑到任务的冲突比例和社会属性是很有必要的。从现实生活来看,形成传统规范的性别角色观念并不容易新的社会规范所打破,以群体为基础的性别刻板印象及偏见可能会扎根于人们的思想难以做出改变。在这样的情况下,减少关注性别与刻板信息之间的联结,而是集中于关注任务中的目标信息,可能为削弱性别刻板印象的表达提供了一种有效的方式。不同的任务冲突比例引发的认知控制对内隐偏见或者刻板印象的表达有一定的调节作用。而不同呈现形式的刺激信息和社会意义之间的转换范式,证明了认知层面的抑制性控制可在特定的呈现形式和跨任务之间减少偏见或刻板印象反应的出现,使得内隐偏见或刻板印象的减少和消除不仅仅局限于刻板印象出现后再进行反应性的加工。

尽管本研究有一些重要的发现,但也存在一些不足。首先,认知控制是目标导向行为的基础,个体是如何根据不同的冲突任务感知对当前目标进行认知控制调整的不得而知,由练习效应而诱发的比例效应和基于目标而引发的比例效应有何区别。这些问题仅借助行为分析并不能使得到有效的回答,也不能保证实验结果的充分性。所以观察不同冲突比例的任务过程中的脑电水平,识别冲突监测和控制调整的脑区差异,有助于我们从客观的角度深入探究冲突水平背后的认知神经机制,为我们在不同的冲突环境下进行有效的认知控制提供理论基础。其次,冲突适应效应和比例效应是认知控制的两种表现形式,冲突适应效应强调个体对先前试次冲突的监测导致在当前试次中更好地控制了冲突,而比例效应则更为关注同一冲突任务类型内监测到不同试次冲突水平的差异引发的冲突适应现象。该研究并没有涉及到两种效应的对比分析,也未深入思考两种效应引发的认知控制有何异同之处。

.....

**编委复审：**建议把题目改为“基于不同冲突水平的认知控制对性别刻板印象表达的影响”。  
并建议作者再仔细通读全文，做最后的文字修改后发表。

**回应：**感谢编委的建议。综合各方面建议，将题目修改为“基于不同冲突水平的认知控制对性别刻板印象表达的影响”，并对文章的语言表达和格式进行了修改。

---

### 第三轮

**主编终审：**经过多轮审修，研究的写作有了实质性的提升。同意发表。

**回应：**非常感谢主编的意见！