

《心理学报》审稿意见与作者回应

题目：比例推理的过度使用及其认知机制：一项发展性的负启动效应

作者：江荣焕 李晓东

第一轮

审稿人 1 意见：

本研究采用抑制控制模型的理论视角和负启动范式探讨了过度使用比例推理的认知机制，研究发现大中小学生在解决一定形式的加法问题时需要抑制比例策略，作者认为这支持了抑制控制模型的观点。作者做了 4 个实验，从研究方法来看较规范，结论较可靠，但存在如下值得商榷的问题。

意见 1：

第 5 页最后一段，“因此，本研究的最后一个目的则是探讨数学问题中的数字比类型（整数比、非整数比）对过度使用比例推理有何影响”。本人认为这似乎是多余的，因为已有研究已经解决，建议不作为本研究的目的之一。

回应：

本研究的最后一个目的是探讨不同的数字比类型对被试在解决加法问题时对比例策略的抑制过程的影响，也就是探讨数字比类型对负启动效应的影响。针对审稿人提出的意见，我们已经修改了引言部分最后一段的表述。

意见 2：

作者的整个研究剔除数据的比例很高，即使是大学生在实验 2 也有 12 人的数据被剔除了，实验 3a 总共 52 名（大学生），居然又剔除了 25 人，几乎占了一半的学生，着实令人怀疑作者的实验结果的可靠性。

回应：

由于过度使用比例推理是一种比较难以克服的倾向，在实验 2 中，大学生剔除的 12 人占总体人数的 27%，我们认为这是一个正常的剔除率，依据 Lim 和 Morera (2010) 的研究（参考文献：Lim, K. H., & Morera, O. (2010). Addressing impulsive disposition: Using non-proportional problems to overcome overgeneralization of proportionality. In *Proceedings of the 13 th Annual Conference Research on Undergraduate Mathematics Education. Raleigh, NC*），国外职前数学老师（*pre-service teachers*）在没有经过干预的情况下，被试进行了前中后三个测试（*pre-test, post-test, end-test*），他们过度使用比例推理的平均比率分别为 41%、57%、25%，即使在有进行干预的情况下，这一比率也达到了 37%、42%、22%（见文章中表 1，

Nonprop MVP (missing-value problem, 非比例缺值问题) 部分的结果)。本研究已经通过呈现合适的指导语、练习和反馈等降低了被试犯错误的倾向, 尽量提高了有效被试的比率。

除了成人的研究之外, 大多数国外的研究都表明小学生到初中生过度使用比例推理的倾向是很严重的, 许多中小学生对比例推理的比率都达到了 50% 以上, 具体可见本文引用的几篇关于中小学生对比例推理的文章。

另外, 关于审稿人提出的实验 3a 中被试问题, 其实实验 3a 中的有效被试是 52 名, 被试总共是 77 名, 剔除的被试是 25 人, 因此剔除率在 30% 左右, 并不是审稿人所说的将近一半。

意见 3:

实验 1 的 8 年级有 38 人 (男 18 人, 女 22 人), 似乎哪里错了?

回应:

已在文中订正。

意见 4:

实验 1 需要描述一下实验设计, 即自变量、因变量是什么?

回应:

已增加每一个实验的实验设计。

意见 5:

实验 1 中探测刺激持续 18000ms (大学生为 15000ms)。为什么要呈现如此长的时间, 而且不同年龄组还不一样? 从后面的结果来看平均反应时为 7 秒左右, 而作者呈现了 18 秒之久, 这可能在时间上被试没有任何压力, 导致负启动量在年龄组之间没有差异? 这也可能是没有在正确率上出现条件主效应的原因?

回应:

首先, 实验 1 的题目呈现时间如此长, 是因为实验 1 采用的是文字应用题, 被试需要一定的阅读时间来理解题意, 而且需要推理才能正确解答, 不是简单的比较, 因此呈现的时间相对要长一些。同时基于负启动范式的逻辑 (启动项目必须正确, 探测项目的数据才有效), 为了保证被试的正确率, 我们在设置时间压力时并不能按照被试的平均反应时来定, 而是要给被试留出充分的时间作出应答。本实验的呈现时间设置一方面参考了以往类似实验范式的研究 [李晓东, 徐雯, 李娜燕. (2012). 潜逻辑运算类皮亚杰守恒任务中的负启动效应. *心理科学*, 35(2), 358-363.] 设置了 3000ms 的时间压力, 但是被试的平均反应时最高也只有 1281ms, 呈现时间是平均反应时的 2 倍多; 另一方面我们也做了预实验以便决定正式实验时的呈现时间。结合两方面的数据, 我们按照平均反应时 2.5 倍的时间设置了时间压力。

其次, 大学生与中小学生的任务呈现时间之所以不一致, 是因为大学生阅读能力和数学

水平比中小学生要高，他们的平均反应时也比6年级和8年级少了1000ms-1500ms，为了给他们造成时间压力，我们相应地减少了呈现时间。通过这样的设置，我们既能保证被试能够有充分的时间作答，保证被试数据的有效性，同时也能够让被试在有时间压力的情况下完成任务（如果停留太久，题目便会消失，而不是直到作出反应才消失，造成时间压力）。

当然，由于以前关于推理任务的负启动研究并没有对刺激呈现的时间压力作出严格的要求，本研究设置的时间压力具有一定的主观性，但是研究的结果也表明，如果依据平均反应时来设置时间压力，那么被试的错误率可能会更高。本研究设置了 18000ms 和 15000ms 的呈现时间的情况下，有些被试依然不能做出正确反应，被试剔除率在 21%到 43%之间，如果进一步缩短呈现时间，可能意味着更高的剔除率，对实验结果造成影响。因此，我们认为这样的时间压力设置是可行的。

最后，审稿人提出任务的呈现时间过长可能导致负启动量或正确率没有年龄差异，我们认为这也是一种解释。例如 Lubin 等人（2013）的研究（Lubin, A., Vidal, J., Lanoë C., Houdé O., & Borst, G. (2013). Inhibitory control is needed for the resolution of arithmetic word problems: A developmental negative priming study. *Journal of Educational Psychology*, 105, 701–708.），要求被试比较简单文字应用题，刺激呈现的时间是不限时的（until response），也就是并没有给被试施加时间压力，结果未发现负启动量有年龄差异。本文在实验 2 的讨论部分补充了这部分内容（见 3.5 讨论，最后一段绿色字体）。

意见 6:

实验 1 刺激呈现流程图这两个流程是完全一样的？

回应:

已对图 1 作出了修改，在控制条件中，启动项目为中立问题。

意见 7:

作者在讨论部分用负启动来解释其研究结果，似乎也可用习惯化或负迁移来解释实验一的结果？

回应:

习惯化是指刺激持续作用后反应强度下降的现象，一般用于婴儿的研究，如婴儿对持续的刺激注视时间会变短，如果刺激发生了一定的变化，婴儿注视时间会增加，表现为去习惯化。本研究中启动项目与探测项目是不同的题目，16 个试次的题目也是不同的，并不是同一题目持续呈现，因而不存在习惯化现象的。测试组在探测项目上的反应时延长也不能解释为去习惯化，因为在控制组探测项目相对于启动项目也是有变化的。如果是去习惯化，测试组与控制组在探测项目上的反应时是不应该有差异的。本实验结果有差异，是因为测试组需要抑制启动项目的误导策略，但是控制组则不需要抑制。

负迁移是指先前的学习对后面的学习产生了消极影响。在本研究中，测试条件下先学习

的是加法问题，后学习的是比例问题；控制条件下，先学习中立问题，后学习比例问题。如果是负迁移，在测试组和控制组都应该在比例问题（探测项）上有较高的错误率，但实际结果是错误率只有 10%左右。另外，如果是负迁移，测试组和控制组在探测项上的反应时也不应该有显著差异，但实验结果是有显著差异的，产生了负启动效应，负启动效应代表着完成任务需要抑制控制的参与。此外，本研究在实验中题目呈现是随机的，试次之间有掩蔽刺激呈现，因此可以排除练习效应的出现。

意见 8:

在实验 2 的讨论部分，作者指出：“本研究也存在类似的结果，从各年级的被试剔除率来看，小学生的剔除率最高，其次为中学生，大学生最低”，但实际结果是小学生只剔除了 9 人，而大学生剔除了 12 人，大学生剔除最多？这似乎是自相矛盾？

回应:

已针对这一问题修改了相应的讨论部分（见 3.5 讨论，第二段绿色字体）。

意见 9:

实验 2 讨论最后一段，“在实验 1 和实验 2 中，变量之间数值的比均为整数，以往有研究表明数字比会影响比例策略的使用，相对于非整数比，整数比诱导出更严重的过度使用比例推理的现象（Fernández et al., 2012; 2011; 李晓东等, 2014）。那么，实验 1 和实验 2 的结果是否是因为整数比所导致的呢？”本人认为前两个实验的共同结果是一种负启动，即干扰效应，怎么与是否为整数比有关系呢？这种推断与假设似乎不成立。

回应:

已经针对这一问题进行了修改（见 3.5 讨论，最后一段绿色字体）。

意见 10:

实验 3a 实验结果部分，“剔除了数据之后，分别对探测项目的错误率和反应时进行 2（实验条件：测试 Vs. 控制） \times 3（年级：6 年级、8 年级、大学生） \times 2（数字比类型：整数比 Vs. 非整数比）”这种表述似乎无法看清实验的自变量及其各自的水平，因为根据 4.1.2 的描述，数字比类型这一变量只发生在加法题目即测试条件中，在控制条件中并不涉及这一变量？

回应:

本研究的主要目的是探讨被试在解决加法问题时是否需要抑制控制的参与，实验中加入数字比类型这一变量也是对加法问题进行操控。由于负启动范式的特性，虽然我们操控了加法问题，但是我们并没有分析加法问题的反应时，而是分析比例问题的反应时，而不同测试条件下的比例问题的反应时又需要进行对比才能得出是否存在负启动效应或者负启动效应是否会受到其他自变量的影响。因此，在负启动范式中，控制条件与实验自变量的操控无关（控制条件不存在抑制），但是实验结果的分析却离不开控制条件下比例题目的反应时，因为控

制条件下比例问题的反应时充当了基线。

此外，如果我们直接将负启动量（两种条件下比例问题的反应时相减）作为因变量进行分析则不存在审稿人提出的问题，但是在进行负启动量的分析之前，我们首先要确保两种条件下比例问题的反应时的差异是显著的。因此，我们依然需要将比例问题的反应时作为一个因变量进行 2（实验条件：测试 Vs. 控制） \times 3（年级：6 年级、8 年级、大学生） \times 2（数字比类型：整数比 Vs. 非整数比）的重复测量方差分析。

审稿人 2 意见：

本研究采用负启动范式研究了过度使用比例推理的认知机制。针对以往研究中存在的不足，本文设计了系列的实验。实验结果表明，被试对于缺值应用题和图片推理题均呈现显著负启动效应，说明被试在问题解决过程中需要抑制比例策略。本研究也讨论了年龄及数字比对比例策略抑制过程的影响。本研究具有一定的理论价值，但仍存在以下一些问题：

意见 1：

摘要过于简略，需要进行相应的扩充。

回应：

已经依据审稿人的要求扩充了摘要（需要指出的是，在投稿指南中，心理学报的要求是中文摘要不超过 200 字，我们也是在此基础上写出了较简略的中文摘要）。

意见 2：

抑制控制模型的引入缺乏逻辑性或论述不足。作者指出，“新皮亚杰学派提出的抑制控制模型则能够弥补上述研究的不足”。实际上，双加工理论框架下，系统 1 和系统 2 的竞争也可以类似地解释比例推理的过度使用。因而此处对双加工机制的评述不够恰当，缺乏说服力。

作者在引述相应的观点的时候缺乏文献引用，如“此外，也有研究者指出对立的双加工系统可能不适用于解释推理过程”。实际上，双加工理论的最新进展也已被用于解释人类推理过程中的信念偏差（belief bias）。因此，不清楚作者基于哪些文献得出上述结论。

回应：

我们在文中补充了 Osman 和 Stavy（2006）文章中的观点，即基于技巧的推理往往包含了启发式加工和分析式加工的特点。例如，被试在解决缺值应用题时，需要对文字进行理解，之后表征问题，还要依据题目中的数字进行计算，当被试过度使用比例推理时，我们很难断定推理过程没有分析式加工的参与，也很难确定比例推理的过度使用是由于启发式“打败”分析式的结果。但是用抑制控制模型来解释则不存在这样的问题，因为在抑制控制模型中，相互竞争的是不同的策略而不是两种对立的加工过程，而不同的策略不存在相互叠加的问题（一个策略不可能既是加法策略又是比例策略）。因此，抑制控制模型更适用于解释比例推

理的过度使用。

第二轮

审稿人 1 意见：

作者较好地对相关问题进行了回复和修改，建议录用当前版本。

审稿人 2 意见：

作者认真地回答了我所提的问题，并根据意见进行了全面的修改。

第三轮

编委专家意见：

建议发表。建议责任作者在发表前务必仔细阅读全文，彻底消除文字及表达上的瑕疵。

回应：

已认真通读全文，修改了部分措辞和错别字（见文中红色字体）并精简了一些表达。