

Theoretical Explanation and Neural Mechanism of the “Hypercorrection” Effect

Wenli Bao, Yuanyuan Xin

School of Psychology, Southwest University, Chongqing
Email: 1129531981@qq.com

Received: Nov. 28th, 2014; revised: Dec. 8th, 2014; accepted: Dec. 17th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Hypercorrection refers to the phenomenon that after the correct answers of questions are fed back to the individual in the test, in the following tests, comparing to the wrong questions of low confidence level, those of high confidence level are more likely to be corrected. There have been three kinds of theories to explain the phenomenon. Attention theory holds the view that people feel surprised when they have failed in questions which they are sure to get the right answers, so, it will arouse their attention and people will spend more time learning the correct answers. Familiarity theory points out that people have known the knowledge of the questions of high confidence level, and learning the information existing in the memory is much easier than learning new knowledge. Recent study area theory tries to explain this phenomenon from a teaching perspective that learning knowledge in recent study area is not difficult, but it will not be extracted directly without relearning (e.g. feedback), so the feedback itself is the key. Hypercorrection's neural mechanism study found that the brain regions of metamemory which does not match are involved in error correction process; especially dorsolateral prefrontal is activated only in the high-confidence but wrong answers, which is the same as right dorsolateral prefrontal cortex. So dorsolateral prefrontal is the inhibition mechanism of false memories. In future research we should pay more attention to the areas of some special groups, such as memory disorder and learning disabilities; this is extremely valuable to know human advanced cognitive process.

Keywords

Hypercorrection, Theoretical Explanation, Neural Mechanism

“矫枉过正”效应的理论解释及其神经机制

包文莉, 辛媛媛

西南大学心理学部，重庆
Email: 1129531981@qq.com

收稿日期：2014年11月28日；修回日期：2014年12月8日；录用日期：2014年12月17日

摘要

在测验中将错误题目的正确答案反馈给个体后，在随后的测验中，信心水平高的错误题目比信心水平低的错误题目更容易被改正，这一现象被称作“矫枉过正”。目前关于“矫枉过正”现象有三种理论解释。注意力理论认为相对于信心水平低时出错，人们发现自身信心水平高的题目竟然会犯错误，因此会唤起注意力，并且花费较多精力去学习正确答案；熟悉性理论指出，那些信心高时所犯错误的一般知识可能是人们一开始就知道的，学习一种已经存在于记忆中的信息远远比学习一个新颖的知识简单；最近学习区理论从学习的角度进行阐述“矫枉过正”可能的心理机制，最近学习区中的知识材料学起来不太困难，但是在没有进行再次学习(如反馈)的情况下也不会被直接提取出来，因此反馈本身很关键。“矫枉过正”的神经机制研究发现在错误矫正时，除了代表元记忆不匹配的共同脑区发生活动，背外侧前额叶只有在高信心回答错误的情况下发生活动，在低信心回答正确时没有出现活动，右背外侧前额叶也是如此，因此可以认为背外侧前额叶参与了错误矫正过程，是错误记忆的抑制机制。未来可以考虑在一些特殊人群等领域展开大量研究，如记忆障碍者、学习障碍者是否存在“矫枉过正”现象，这对于了解人类高级认知过程极具价值。

关键词

矫枉过正，理论解释，神经机制

1. 引言

“失败乃成功之母”，毫无疑问，为了做出更加合理有效的决策，人们总要不断的改正错误。比如学生的考试、商业投资、政策法规的制定，我们所知道的任何一种人类活动都会经历错误矫正过程。可以说，科学研究本身就是一个发现错误改正错误的过程，因此深入地了解错误矫正过程对探究不断试误和更新的人类认知过程具有关键作用。

前人研究发现，当个体犯错误之后，给其反馈正确答案往往可以有效地使错误得到纠正(Anderson, Kulhavy, & Andre, 1972; Butler, Karpicke, & Roediger III, 2008; Butterfield & Mangels, 2003) (Kang, McDermott, & Roediger III, 2007) (Metcalf & Kornell, 2007; Metcalf, Kornell, & Finn, 2009)，这一结果与我们的一般观念也是相符的。但 Butterfield 和 Mangels (Butterfield & Mangels, 2003)的研究指出，将正确答案反馈给个体这一方法对随后测验成绩的作用大小，还与个体犯错误时的信心水平有密切的关系：人们犯错误时信心水平越高，在后面的测验中错误越容易改正。人们在信心水平较高时犯下的错误往往比信心水平低时犯下的错误更容易改正，这就是“矫枉过正”效应(Hypercorrection Effect) (Butterfield & Metcalf, 2001)。很明显，这里是直译过来的，实际与中文的矫枉过正意思不一样，中文的矫枉过正是指把弯的东西扳正，又歪到了另一边，比喻纠正错误超过了应有的限度。

大量研究已经证明人们对自身反应的元认知通常是准确的(Kornell & Metcalf, 2006; Metcalf et al., 2009)，比如，个体对自己认为正确的事情有较高的信心，反过来，自己非常有信心的事情通常实际上也是正确的。但目前错误矫正领域的相关研究告诉我们信心高的错误更容易改正(“矫枉过正”)，

如果实际情况的确如此，那似乎意味着人们可以很轻易地改变内心坚定无比的信念——这与我们的普遍看法恰恰相反。虽然“矫正过正”现象与我们的直觉相反，但它的确是生活中普遍存在的一种行为，个体总是在不断地犯错误和改正错误中成长，因此关于这一特定错误矫正现象的研究值得重视。

2. “矫正过正”现象的研究范式

目前，研究者没有系统地提出研究矫正过正现象的经典范式，但总结前人研究可以知道一般程序如下。首先，要求被试对一定数量的一般性知识问题逐一作答，每完成一道题目就对答案的正确程度进行信心水平的评估(对自己回答正确的信心水平从“非常确定 0.00”到“非常不确定 1.00”进行评估)；然后，所有题目全部完成之后(这里的“完成”通常控制在被试回答错误的题目达到一定数量的时候，因为错误的题目是研究“矫正过正”现象所需要的材料)，主试将正确答案呈现给被试，被试可以同时看到自己的答案和正确答案；最后，主试会挑选出被试回答错误的题目，再次让被试进行作答(目前有要求被试直接产生答案，多重选择，字母提示(启发)三种方式)并再次进行信心水平的评估。在最后的测验中，比较信心水平低时犯的错误的和对信心水平很高时犯下的错误(信心水平高低可以使用 0.5 为界限，也可以使用一个被试所有题目信心的中值或平均值为界限)，在第二次作答时人们在哪种情况下更容易作对，即分析错误题目的信心水平与再测验成绩之间的关系。在 fMRI 实验中，与进行行为实验主要不同的一点是，被试在扫描之前答题(因为研究者需要挑选被试错误的题目以及得到被试的平均信心水平)，在扫描的时候进行反馈，扫描之后在扫描仪外再次测验；其他步骤基本相似。

一般性知识问题的介绍：Nelson 和 Narens 建立了一个一般性知识问题的题库(题目类型如“测量六英尺水深度的单位名称是什么？(答案：英寻)”或“作品《陌生人》的法国作者是谁？”)(Nelson & Narens, 1980)，其中每道题的正确反应的参考值已经存在，研究者可以从中随机抽取题目作为实验材料。

3. “矫正过正”现象的理论解释

对于矫正过正现象，目前还没有比较成熟和系统的理论模型。从前人研究来看，主要有下面三种观点：注意力说(enhanced attention)，熟悉性说(familiarity)和整合说。

3.1. 注意力说

这一观点从注意力的角度出发，当人们发现自己信心很高却回答错误的时候，他们会感到吃惊甚至尴尬，因此他们会集中注意力，全神贯注学习正确答案。一些研究支持了这一观点，比如 Butterfield 和 Metcalfe 发现在一个同时进行的任务中，当对高信心错误进行正确答案的视觉反馈时，相对于低信心情况，人们更容易忘记去做一个探测软音调(soft tone)的实验任务。这在一定程度上可以说明，低信心下犯的错误的在被矫正时所消耗的认知资源没有高信心情况消耗的多(Butterfield & Metcalfe, 2006)。

在该问题上还有 Erp 研究，Butterfield 和 Mangels 发现被试在出现“矫正过正”效应的情况中，得到正确答案的反馈时，脑电数据会出现一个 P300 波，而且高信心错误收到反馈时比信心低错误得到反馈时的 P300 更大(Butterfield & Mangels, 2003)。前人的研究已经说明 P300 表示的是对新异刺激的强烈注意，经常在 oddball 范式的实验过程中被记录到，所以虽不能确定因果关系，但一定程度上可以认为“矫正过正”与注意力关系密切。

另外，Fazio 和 Marsh 也发现，情景方面的记忆更容易形成和保持(比如说人们往往能够把表面状况(surface appearance)记得更牢固)，也能说明对高信心时所犯错误的改正这种记忆被加强了，同时也加强了对这些错误的正确答案的注意和记忆(Fazio & Marsh, 2009)。

3.2. 熟悉性说

第二种解释叫做“熟悉性解释”，可以理解为被试对那些高信心错误的题目或答案其实相对更熟悉一些。形象的说法是“我从一开始就知道(答案)是这样(know it all along)”。该说法大概思想就是，那些一般性知识问题的特征和那些信心水平高时所给出的错误答案之间有些系统差异，或者是在参与者错误时，信心高的知识领域与信心低的知识领域，可能存在基于个体的熟悉性上的不同。虽然犯错，但因为对那些信心高的题目比那些信心低的题目更加熟悉，那么在高信心错误领域，就是更熟悉地学习了正确答案(在已有相关知识的基础上进行了再次学习)，而在低信心错误领域，学到的很少，甚至根本没有学习到(第一次学习)。拿一个典型的例子说明，比如“加拿大的首都是__”，有人回答“多伦多”，他的答案就是错误的。不过当该参与者知道正确答案是“渥太华”时，他或许发现这个正确答案很容易就记住了，因为他本来就知道渥太华是加拿大的一个城市，甚至也许本来就知道渥太华是加拿大的首都。他们或许早就知道正确答案，只是在回答的时候不小心犯了错误；换句话说就是说，“越熟悉反而越容易出错”。现在反过来考虑一下低信心时犯的的错误，比如说“巴马科市是布迪的首都”，参与者对“布琼布拉”根本不熟悉(甚至对布隆迪也很少听闻)，所以需要重新学习——他们一开始就不知道是这样。

支持“熟悉性”解释观点的证据也有很多。Butterfield 和 Metcalfe 重新分析了他们 2001 年研究所得的数据(Butterfield & Metcalfe, 2006)。他们 2001 年用的一般性知识问题来自 1980 年 Nelson 和 Naren 的文章，这些一般性知识问题每道题的正确反应的参考值已经存在，因此研究者可以计算各信心水平下犯错题目的正确反映参考概率。结果发现，信心水平越高，该值越高。将每一道题目的特征作为信心水平的函数，差异显著。第一次回答错误题目的标准 ease 与第二次正确回忆的成绩关系显著。当排除标准难度之后，仍然有“矫枉过正”效应。所以说，虽然熟悉性并不是“矫枉过正”的原委，但它起了相当大的作用。

Butterfield 和 Mangels 要求他们的被试评估每道题目的主观熟悉性。结果发现，被试觉得自己对那些信心高犯错的题目比信心低犯错的题目更熟悉。而且他们发现，在事件相关电位的数据中分析得，一个前额叶负波出现在正确答案呈现之后的 300~600 ms 的时间区间内，但这种情况也只有在答案包含熟悉的语义信息时才有(Butterfield & Mangels, 2003)。这个负波在一定程度上说明了题目与先前存在的语义信息之间是有联系的。换句话说，那些信心高犯错的题目或题目与答案可能已经在语义记忆中存在了。

但对题目的熟悉并不能说明个体在一开始真的知道答案。因为在实际实验操作中是将答案反馈给被试之后，被试发现错误才会有“我早就知道”的反应，这很像生活中一种常见的现象——“事后诸葛亮”(Hindsight Effect)。熟悉性没有将“只熟悉题目本身”和“熟悉题目和题目的答案”进行区分，要想进一步了解“矫枉过正”，探究被试是否真的知道答案也很有必要，这方面既有来自成人的研究，也有来自儿童的实验结果。Metcalfe 和 Finn 发现成人和儿童都会出现过分矫正现象，但不同的是，再次测验时，无论多重选择回答方式还是被试直接产生答案的方式，对那些信心高时犯的的错误，成人的表现很好，而儿童往往不能直接作答，甚至不能在多个选项中正确选择(Metcalfe & Finn, 2012)。究其原因，这可能跟儿童的生理发育状况有关。比如儿童不像成人那样能直接产生答案，可能是因为儿童的语言流畅性较差。因此从成人的研究结果来看，他们的确“从一开始就知道了”，即熟悉答案。

3.3. “最近学习区”解释说

作者认为，由于“熟悉”的定义很宽泛，研究者很难统一测量熟悉程度，因此熟悉说还欠缺科学性，而最近学习区解释说是从学习的角度进行阐述“矫枉过正”可能的心理机制，也可以认为一种是对注意力说和熟悉说的整合性解释。

Metcalfe 和 Finn 认为，之所以在高信心时犯的的错误比低信心时犯的的错误更容易改正，可能是因为高

信心错误已经存在于人们的最近学习区(person's region of proximal learning, RPL) (Metcalfe & Finn, 2011)。最近学习区是指具有那些几乎学习过但同时没有完整学习过的材料, 这种情况下个体一旦获得学习机会就能获得或记忆相应的知识。一般认为, 最近学习区中的知识材料学起来不能太困难, 但是在没有进行再次学习(如反馈)的情况下也不会被直接提取出来。Metcalfe 和 Finn 用实验验证了这一点, 在没有反馈的条件下(没有学习机会), 被试那些高信心错误的再测成绩与初试成绩没有差异, 可见错误不会自动矫正; 而给予反馈的条件下, 被试的再测成绩显著提高, 并且高信心错误的的成绩比低信心错误的的成绩更好。因此反馈本身很关键, 而高信心错误之所以容易矫正的究其原因, 可能既包含对正确答案的注意力加强(学习过程需要注意力), 也包含人们真的一开始就熟悉题目答案这一原因。

4. “矫正过正”的神经机制

目前关于“矫正过正”神经机制研究的数量很少, 从已有研究来看, 研究者主要从元认知不匹配(具体是元记忆不匹配)的角度, 通过对比“高信心回答错误”和“低信心回答正确”这两种情况, 探究和解释“矫正过正”现象。

Metcalfe, Butterfield, Habeck 和 Stern 首次对矫正过正现象进行了功能性磁共振研究。他们假设, 从元认知的角度来说, 高信心与回答错误, 低信心与回答正确都是不匹配的, 这一不匹配的意识反应会出现在被试接收到正确答案时。经过功能性磁共振研究, 结果发现, 首先, 无论高信心时回答错误, 还是低信心时回答正确, 一些与元记忆(metamemory)不匹配有关系区域的大脑活动都会出现, 其中包括横跨双侧前扣带回的额内侧, 双侧额内侧回和右扣带(Metcalfe, Butterfield, Habeck, & Stern, 2012)。这些区域的活动都可以说明, “矫正过正”效应一定程度上来源于元认知冲突过程。这与上文“矫正过正”的ERP的研究结果也基本是一致的。

而对“矫正过正”本身更重要的结果是, 研究发现除了说明元记忆不匹配的共同脑区(主要是前扣带和相关的前额皮层)活动之外, 还有两个例外脑区只在犯错误时被激活。其中, 背外侧前额叶只有在高信心回答错误的情况下发生活动, 在低信心回答正确时没有出现活动。而右背外侧前额叶也是如此。因此, 我们可以认为背外侧前额叶参与了错误矫正过程, 是错误记忆的抑制机制。在抑制错误答案的同时, 也加强对正确答案的注意, 这一研究揭示了“矫正过正”效应可能的神经机制。

5. 研究展望

虽然研究者从各个方面对“矫正过正”现象展开了一些研究, 但数量还比较少, 研究也不够成熟和系统, 目前还有许多尚未解决的问题, 今后可以着重从以下的几个方面进行研究。

第一, 正常人犯错误经过学习正确答案之后可以使错误顺利得到矫正, 而记忆障碍者或学习障碍者快速地改正错误有一定困难(Glisky, Schacter, & Tulving, 1986; Wilson, Baddeley, Evans, & Shiel, 1994), 因此错误对他们更加不利。如果对这些特殊人群的一般信心水平与错误矫正的关系进行考察, 可以帮助我们逐步完善目前的结论并发现应用价值, 因此未来很有必要选择特殊人群继续深入研究。

第二, “矫正过正”的脑机制研究还在刚刚开始阶段, 我们只得到背外侧前额叶区域可能参与这一过程的结果。由于“矫正过正”主要涉及注意和记忆过程, 也许可以从与记忆、注意相关的脑网络进行考察, 逐渐找出该行为的神经活动模式, 使矫正过正的脑机制越来越清晰。

第三, “矫正过正”的适应意义也是今后值得研究的一个重要方向。本来可以说“越坚定越不容易被摧毁”, 但按照矫正过正的说法就变成“我们越坚定的信念一旦遭遇否定, 反倒轻易就会改变”。这一现象既然已经存在, 那么它必有道理, 因此可以从进化和适应的角度考察“矫正过正”在生活中可以给人带来哪些优势。

参考文献 (References)

- Anderson, R. C., Kulhavy, R. W., & Andre, T. (1972). Conditions under which feedback facilitates learning from programmed lessons. *Journal of Educational Psychology*, *63*, 186.
- Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 918.
- Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research*, *17*, 793-817.
- Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1491.
- Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, *1*, 69-84.
- Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, *16*, 88-92.
- Glisky, E. L., Schacter, D. L., & Tulving, E. (1986). Learning and retention of computer-related vocabulary in memory-impaired patients: Method of vanishing cues. *Journal of Clinical and Experimental Neuropsychology*, *8*, 292-312.
- Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528-558.
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 609.
- Metcalfe, J., & Finn, B. (2011). People's hypercorrection of high-confidence errors: Did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 437.
- Metcalfe, J., & Finn, B. (2012). Hypercorrection of high confidence errors in children. *Learning and Instruction*, *22*, 253-261.
- Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*, 225-229.
- Metcalfe, J., Butterfield, B., Habeck, C., & Stern, Y. (2012). Neural correlates of people's hypercorrection of their false beliefs. *Journal of Cognitive Neuroscience*, *24*, 1571-1583.
- Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, *37*, 1077-1087.
- Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, *19*, 338-368.
- Wilson, B. A., Baddeley, A., Evans, J., & Shiel, A. (1994). Errorless learning in the rehabilitation of memory impaired people. *Neuropsychological Rehabilitation*, *4*, 307-326.