

《心理学报》审稿意见与作者回应

题目：面孔熟悉过程中部件加工与整体加工的区域异步性和眼睛区域特异性

作者：王哲;倪昊;封丹;严璘璘;孙宇浩

第一轮

审稿人 1 意见：

本研究探究了接触 1 个月和接触 13 个月的两组大学生对同学面孔和陌生人面孔的识别绩效，发现接触 13 个月组的大学生表现出更强的熟悉度效应。此外，针对完整面孔、区域部件和区域外周条件的分析，得出熟悉过程中眼睛区域的整体加工能力先于部件加工能力在识别中发挥作用，提示面孔熟悉过程具有眼睛区域特异性。本研究对于理解经验对面孔加工的塑造具有一定的意义。

意见 1：本研究主要有两个重要问题需要进一步明确。

第一，面孔再认上的差异是不是由于经验/熟悉度不同导致值得商榷。因为作者并不是随机分配被试到两个组别中，这导致很多混淆因素无法控制和排除。此外，两组被试看的熟悉人脸是不一样的，那么是不是由于刺激的差异导致再认绩效不一样？

回应：审稿人的意见，在此处包含两点。下面逐一回复。

第一点是关于被试分组。本研究没有采用随机分配法，有可能造成变量混淆。对这一点，解说回复如下：

本研究对被试分组的方法确实并非随机分配法，而是招募某大学某学科的大一年级的整个班（相互接触 1 个月）和大二年级的整个班（相互接触 13 个月）分别作为“低熟悉组”和“高熟悉组”。在《心理学研究方法》中，这种方法被称为“横断设计”（也称为“横向设计”）。

横断设计并非“真实实验”设计，而是一个“准实验”设计。在发展心理学研究中，它是最常用的三种研究设计之一（注：另外两种是纵向设计和聚合交叉设计）。例如，以某小学一年级、二年级、三年级的学生为三组被试，横向比较三者任务中的“平均绩效”或“效应模式”。

在认知心理学中，专家技能精通（Expertise）领域的研究也常用横断设计作为“准实验”设计。例如，以国际象棋的新手、高手和大师为三组被试，横向比较三者完成棋局认知任务的“平均绩效”或“效应模式”。

横断设计的优点是代表性好，研究结果的外部效度较高，并且可以避免纵向设计的重复测量带来的练习效应。横断设计的主要弱点在于，因为被试群体不同，所以被试群体本身的某些差异可能会造成实验结果的混淆（例如，同辈效应[cohort effect]，两组被试因其经历社会时期的不同而出现与年龄变化伴随的差异）。

我们认为，本研究采用横断设计是合适的。当然被试群体本身的差异有可能造成混淆，但我们认为它是微弱的、不影响本研究结论的。原因主要包括 4 个方面。（1）两个年组的被试是同一所大学同一学科，所以就人口学变量而言，可以认为他们有良好的组间同质性。（2）因为两组被试都已经是成年人而且年龄只相差一岁，可以认为社会时期变化和“同辈效应”非常微弱，或是没有。（3）两组被试最可能具有的差别，是普遍性的。例如，刚刚入学 1

个月和已经入学 13 个月的被试对大学生活的适应性可能有所不同；他们对心理学实验的认识、反应性和情绪波动可能不同。这类普遍性的差异有可能会影响“平均绩效”，但不会影响“效应模式”，因此，本研究的核心发现不会造成严重的混淆。例如，大一新生在他们的所有实验条件中的“平均绩效”可能会普遍比较低；相对来说，大二学生的“平均绩效”可能会普遍比较高。但是，本研究的理论预判和实验发现，主要聚焦于每个年级的“效应模式”是什么样，有何差别，有何共性——我们特别关心被试在“实验任务”和“面孔区域”这两个自变量上（每一个年级的结果）**分别**有何种主效应（例如，眼睛和嘴巴的区域差别，三种加工方式的**任务差别**）、交互效应和简单效应。两组被试的普遍差别所可能造成“平均绩效”高低的混淆，但不会影响到这些“效应模式”。实际上，两个年级的平均绩效有差异，主要是显示出熟悉度的主效应。而就目前所得的结果来看，我们相信熟悉度效应在两个年级的“平均绩效”之间是真实存在的。（4）我们给两组被试呈现的“陌生面孔”是相同的，相应的平均绩效是稳定的。可以从中看到两组被试有没有差别、有多大差别。综上，我们认为，本研究采用横断设计是合适的，不会造成严重的混淆。

第二点是关于实验材料，“两组被试看的熟悉人脸是不一样的”。解说回复如下：

本研究给两个年级的被试分别呈现了他们各自“熟悉”的面孔。因为两组被试真实的同班同学确实不一样，所以二者的“熟悉面孔”也不一样。这是本研究在实验材料方面相对于常见的实验设计所不同的地方。（注：给两组被试呈现的“陌生面孔”是相同的，平均绩效是稳定的。）

我们曾经考虑，如果“两组被试看同一组的熟悉人脸”（例如，既然是同一学科，用讲课教师的面孔照片，两组学生也是分别熟悉了 1 个月和 13 个月），是否可行？如果这样做了，那么本研究的内部效度一定会提高。然而，考虑到四个原因（理论上 1 个原因、方法学 1 个原因、现实中 2 个限制），最终我们还是采用了目前这种方式来设定实验材料。

理论上的 1 个原因是，本研究界定的“熟悉脸”是在真实生活中密集、多次接触的真人面孔。同班同学可以达到这种程度，但是讲课教师（即使经历了 13 个月），也很可能仍然达不到这种程度。方法学的 1 个原因是，从真实人群和真实熟悉得出的实验结果，会大幅度提高本研究的外部效度，有现实意义。现实中的 2 个限制在于：（1）“既要真实，又要能够获得，还要有足够多的数量”，同时满足上述三条要求，这样群体的被试和群体的照片，最现实的就是同班同学；（2）陌生人的大脸照片（“大头照”）看起来比较“不美”（一些被试在实验后做如此的报告），但同班同学的面孔照片却似乎不受这种妨碍。

更重要的（与第一点回复的第 3 方面原因相似），即使实验材料的差别会影响两个群体“平均绩效”，也不会损伤本研究所聚焦的理论假设和核心实验发现。因为（1）面孔熟悉度效应在以往研究中多次重现，稳健性（robustness）较好；（2）本研究只“定性”预判大一新生和大二学生的“平均绩效”有显著差异，不“定量”考察（也无需关心）二者有具体多大差异；（3）我们特别关心两组被试在“实验任务”和“面孔区域”这两个自变量上（每个年级的结果）**分别**有何种主效应、交互效应和简单效应。而两组熟悉面孔作为实验材料的差别只会造成“平均绩效”的提高或降低，不会影响到“效应模式”。

关于上述两点的回复，在修改稿的[前言]、[方法]、[结果]和[讨论]中都有增添与修改（请见第 1 节第 7 段和 8 段、第 2.1 节第 1 段、第 2.4.1 第 2 段和第 3 节第 1 段的蓝字）。综合起来，我们认为本研究采用横断设计分配被试和两组熟悉面孔作为材料是可接受的，能满足严

谨性的要求。当然，实验刺激本身的差异应当得到重视，以及感谢审稿专家指出横断设计本身存在的弱点。

意见 2: 作者研究的到底是不是面孔再认也值得商榷。由于学习阶段和再认阶段，同一面孔完全一样，如何认为本研究是在研究面孔再认而非图片匹配？此外，从图 2 中可以看出不同面孔刺激的低水平特征（比如亮度、颜色）没有匹配，可否认为被试在匹配低水平特征，而非面孔身份？

回应: 感谢审稿人指出这一点。面孔学习和面孔再认阶段，同一人（不论陌生还是熟悉）采用了同一样的正面照片。存在一种可能性，即图片匹配（低水平视觉信息加工）和面孔再认（高水平视觉信息加工）都有发生。换言之，在再认阶段，低水平视觉信息加工和高水平信息加工二者可能兼有。同时我们认为，不存在另一种可能性，即“只发生了图片匹配而完全没发生面孔再认”。理由有三。

（一）面孔加工会在很短时间内自动化进行，行为（视觉搜索）和脑电证据（例如，ERP 的 N170 波）都提示，被试很难自上而下地随意抑制面孔加工。而且，陌生面孔是如此，熟悉面孔加工的自动化程度更高（例见 Wiese, H., Anderson, D., Beierholm, U., Tüttenberg, S. C., Young, A. W., & Burton, A. M. (2022). Detecting a viewer's familiarity with a face: Evidence from event-related brain potentials and classifier analyses. *Psychophysiology*, 59(1). <https://doi.org/10.1111/psyp.13950>）。

（二）在以往的多个实验研究中，对比过“正面呈现”和“侧向旋转”的面孔照片作为再认阶段材料时被试的绩效，大趋势是面孔侧向旋转导致被试的整体绩效下降（i.e., 出现了一个主效应），但是“正面呈现”和“侧向旋转”的面孔照片结果模式是相似甚至相同的。这就是说，再一次地，其“平均绩效”有所上升，但“结果模式”没有受到影响。换言之，图片匹配（低水平视觉信息加工）确实起到了一些作用，但面孔再认（高水平视觉信息加工）确实发生了，并且在“结果模式”中表现出它会产生怎样的效应。

（三）在本研究中，这些人脸照片因其真实，本身就存在亮度和颜色的差异。我们经过考虑，在三种实验任务和所有实验条件中都使用了同一种方式采集的真实人脸。这种做法很可能导致“平均绩效”（特别是陌生脸）上升，但两种面孔自身的“结果模式”及二者的差异仍然是清晰的。综上，我们相信本研究最关心的效应模式（两种加工的区域异步性和眼睛区域特异性）并未受到严重影响，是可接受的。

关于上述回复，在修改稿的[方法]、[结果]和[讨论]中有增添与修改（请见第 2.2 节第 1 段、第 2.4.1 节第 2 段、第 3 节第 1 段的蓝字）。

意见 3: 作者提到面孔熟悉是一个长期的过程，但面孔熟悉完全可以是一个短期的过程。在大学生活中，同学之间有可能接触不多、接触质量低导致需要长期接触才能达到最熟悉的状态；但如果强迫学生天天进行高质量的接触（比如一起玩耍、做小组项目），可能几天就能达到熟悉性的饱和状态。13 个月组比 1 个月组再认熟悉面孔更优是由于知觉经验更丰富还是他们有更佳的接触质量（比如成为朋友等）导致，这值得讨论中讨论。

回应: 审稿人提出了一种理论上的可能性，我们完全赞同这种可能性值得讨论。修改稿在讨论的第一个发现处，对此可能性做了一些分析和探讨（请见第 3 节第 2 段的蓝色字）。

意见 4: 作者提到“在当前研究中，我们使用同一类的熟悉面孔（同班同学的面孔）控制了

社会经验可能带来的面孔识别差异，结果表明，单纯知觉接触经验的增加就可以提高人们对熟悉面孔的识别能力”，为什么作者会认为只是单纯知觉接触经验导致？

回应：此处“单纯知觉接触经验”的措辞不准确。结合您上一个意见进行了一些分析和新的探讨，现在修改稿删去了这一不准确的观点，改为接触经验和更操作化的接触时间、接触次数和社交经验等变量名（请见第 3 节第 2 段的蓝色字）。

意见 5：学习区域外周的任务为什么能反应面孔整体加工？作者需要阐明下。

回应：区域外周识别任务体现面孔整体加工的逻辑如下：遮挡某一部件（例如，遮挡眼睛），迫使被试在第一个注视点时，只能用外周视野对面孔部件区域的外周区域做加工。因此被试做面孔匹配的绩效可以反映出“面孔知觉场(perceptual field)”的范围（注视面孔某一区域时，面孔其他区域上有多大参与到此感知之中），以此，可以用区域外周任务反映出面孔整体加工程度的高低（也见 Rossion, 2008, 2009; Van Belle et al., 2010）。

关于此点，在修改稿的[前言]（倒数第二段）和[讨论]（第二和第三个发现处）有增补与修改（请见第 1 节第 7 段和第 3 节第 1 段的蓝字）。

意见 6：样本量的确定需要写在正文中。

回应：已在修改稿的被试部分添加相关内容：“G-Power (v3.1) 设定 $\alpha = 0.05$, $1 - \beta = 0.8$, 发现三因素混合实验设计中 2（被试组别：接触 1 个月组、接触 13 个月组） \times 2（面孔类型：熟悉、陌生）的交互效应显著，至少需要被试数量为 34 人，效应量会大于 0.25。”（请见第 2.1 节第 2 段的蓝字）。

意见 7：相关分析有没有多重比较校准？

回应：感谢指出，在原稿中相关分析时并未进行多重比较校准。现在修改稿用 FDR 方法将相关分析的 p 值进行校准（请见第 2.4.4 节第 1、2 段的蓝字）。

意见 8：针对面孔区域的部件加工与面孔识别之间的相互关系，作者有没有对比两组被试的相关系数的差异？

回应：已在面孔区域的部件加工与面孔识别相关分析部分添加两组被试相关系数的差异比较（请见 2.4.4 节第 1 段的蓝字）。

意见 9：“这可能是因为日常社交生活中人们的熟悉过程貌似迅速”。语句不通顺。

回应：感谢指出。修改稿将此句（及其前后文）改为“在真实生活中的社交场景里，人们常常会经历一种‘似乎迅速’的熟悉过程：原先是陌生人，经过简单的互相介绍之后，人们就默认彼此已经处于‘熟悉’的社交状态之中”。请见前言第二段（第 1 节第 2 段）的蓝字。

意见 10：“熟悉面孔的识别绩效会随着接触时间的增加而持续提高”。本文多次提到“持续”这个词，但作者只比较了接触 1 个月和接触 13 个月的差异，这个词的使用有点不妥。

回应：感谢指出，原文确实不妥。修改稿将全文中的“持续”都改为“累积”（accumulated）。这个词与“突变”或“突现”（emergent）相对。

.....

审稿人 2 意见:

意见 1: 陌生面孔是如何选取的? 为何陌生面孔有接触 1 个月和接触 13 个月两种条件? 为何从结果表的 95%CI 数据来看, 陌生面孔的接触 1 个月和接触 13 个月有显著差异?

回应: 原稿此处文字表达不准确。应该为“接触同学 1 个月的被试组(大一新生)和接触同学 13 个月的被试组(大二学生)”。这是两组被试, 对同一批陌生面孔照片做任务。而并非陌生面孔有接触 1 个月和接触 13 个月的两种条件。

修改稿在表格和文字中, 将这两个水平的名称改为“接触 1 个月的被试组”和“接触 13 个月的被试组”, 以此界定清楚是两组被试, 避免误解。

关于 95%CI, 接触 1 个月的被试组和接触 13 个月的被试组识别陌生面孔的绩效差异只反映在反应时上(二者正确率无显著差异, $F(1,55) = 1.960, p = 0.167$)。反应时数据显示“接触 1 个月的被试组”反应较慢(反应时较长), “接触 13 个月的被试组”反应较快(反应时较短), 二者差异显著。出现这种情况, 我们认为有三种可能的原因:(1) 因为在大学的环境中, 从大一到大二, 学生们与人接触的时间、次数和社交经验会逐渐累积。这可能导致大一新生识别面孔的普遍能力相对较弱, 大二学生识别面孔的普遍能力相对增强, 所以二者对陌生面孔的反应时有差异。(2) 因为两个年级对熟悉面孔的熟悉程度有较大差别, 导致在整个实验中, 大一新生做再认判断(既包括肯定反应也包括否定反应)的普遍较慢, 大二学生普遍较快。(3) 可能因为上述两个因素都在起作用。

我们认为, 上述差别对本研究的结果、发现与结论的影响很小。数据显示, 大一新生在他们的所有实验条件中的“平均绩效”会普遍较低; 相对来说, 大二学生的“平均绩效”会普遍比较高。但是, 本研究的理论假设和实验发现的焦点, 并不是比较两个被试组的“平均绩效”, 而是把两个被试组的“效应模式”提取出来, 再做对比。两个年级的“效应模式”分别是什么样, 有何差别, 有何共性——我们特别关心被试在“实验任务”和“面孔区域”这两个自变量上(每一个年级的结果)分别有何种主效应(例如, 眼睛和嘴巴的区域差别, 三种加工方式的任務差别)、交互效应和简单效应。这些“效应模式”不会受到两组被试在“平均绩效”上相对高低的影响。(在我们对审稿专家 1 的问题 1 的回答中, 也做了相近的阐释)。

修改稿在[方法]、[结果]和[讨论]有增补与修改, 请见第 2.3 节 1 段、第 2.4.1 节第 1、2 和 3 段、第 2.4.2 节第 1、2 和 3 段、第 2.4.3 节第 1 和 3 段、第 2.4.4 节第 1 和 3 段、第 3 节第 1 段、2 和 5 段的蓝字。

意见 2: 为何区域外周面孔只遮挡了一只眼睛? 为何区域外周识别任务体现了整体加工? 加入这个任务背后的逻辑是什么?

回应: 区域外周识别任务体现面孔整体加工的逻辑如下: 遮挡某一部件(例如, 遮挡眼睛), 迫使被试在第一个注视点时, 只能用外周视野对面孔部件区域的外周区域做加工。因此被试做面孔匹配的绩效可以反映出“面孔知觉场(perceptual field)”的范围(注视面孔某一区域时, 面孔其他区域上有多大参与到此感知之中), 以此, 可以用区域外周任务反映出面孔整体加工程度的高低(也见 Rossion, 2008, 2009; Van Belle et al., 2010)。

此外, 关于区域外周条件只遮挡一只眼睛或嘴巴, 与“知觉场范式”设定注视点在一只眼睛或嘴巴上, 是保持一致的。我们认为, 这种设定的合理性除了范式自身的上述逻辑之外,

还有三条原因。(1) 在两种区域条件下, 采用了相同的椭圆进行遮挡, 以保证不同条件下面孔信息的损失相同; (2) 如果只使用一个遮挡物, 遮挡两只眼睛的面积势必较大, 使用这么大面积去遮挡嘴巴区域时, 会遮挡住嘴巴之外的大量区域 (比如部分的鼻子区域); (3) 如果用两个遮挡物去遮挡两只眼睛, 就可能引入遮挡数量这个混淆变量。所以, 采取只遮挡一只眼睛的方式, 有利于控制无关变量。

修改稿的[前言] (倒数第二段)、[方法]和[讨论] (第二和第三个发现处) 有增补与修改 (请见第 1 节第 3 段、第 2.3 节第 1 段、第 3 节第 1 段的蓝字)。

意见 3: 最好在设计与流程部分更明显地指出面孔区域这个自变量是通过将该区域呈现在注视点的位置来操纵的, 以及这样操纵背后的逻辑。

回应: 修改稿添加了说明。“操纵自变量 (面孔区域: 眼睛、嘴巴) 的具体技术是, 改变学习刺激在屏幕上呈现的位置, 使眼睛区域或嘴巴区域呈现在屏幕中央的注视点位置, 迫使被试的第一个注视点投放在眼睛区域或嘴巴区域上。” (请见第 2.3 节第 1 段的蓝字)。

意见 4: 结果 2.4.2 中, 为何能从注视被遮挡的眼睛或嘴巴位置的区域识别任务 (实际上判断的是未被遮挡的部分) 的结果就得出以下结论“上述结果表明, 接触时间对不同面孔区域特征识别的作用是不同的, 表现出不同的趋势, 眼睛区域的特征识别正确率随接触时间增加而表现出上升趋势, 而嘴巴区域的特征识别正确率随接触时间增加而表现出先升后降趋势。”

回应: 我们检查了原文, 结果 2.4.2 应该是有关区域识别任务的结果, 识别“区域部件”, 而不是识别“区域外周” (遮挡了眼睛或嘴巴的条件)。结果表明“人际接触时间对眼睛和嘴巴的识别产生的作用有区域选择性, 眼睛区域识别的正确率随接触时间增加而上升, 嘴巴区域识别的正确率随接触时间增加而先升后降。”请见区域部件识别任务第二段 (第 2.4.2 节第 2 段)。

意见 5: 本研究只操纵了 2 个水平的接触时间, 就得出“随着接触时间的增加而持续提高”的结论似乎不妥。2 个水平的比较体现不出“持续”。

回应: 赞同审稿人意见。修改稿将“持续”全都改为“累积” (accumulated), 表示“在一个过程中逐渐增加”的含义。

意见 6: “Baker 等人的研究也有类似的发现, 他们采用拍摄于不同时间跨度内的视频作为实验材料, 发现那些观看了仅在一日内拍摄的视频的 6-13 岁儿童在后续面孔识别中的表现要差于那些观看拍摄了三天以上视频的儿童 (Baker et al., 2017)。”在这个研究中, 被试看视频的时间一样吗? 如果时间一样的话, 不能说被试和视频中的人物的“接触时间越多”。视频内容是一天内的, 还是三天内的, 体现的是 within-person variability。

回应: 被试观看的视频时间是一样的, “接触时间越多”的表达确实存在不妥, 在修改稿中将“接触时间”修改为了“变异性”, 对该部分文字具体的修改为: “Baker 等人 (2017) 让两组儿童分别观看高变异性视频 (同一个人的发型、服装和妆容发生了改变) 和低变异性视频 (同一个人的发型、服装和妆容未发生改变), 发现前者在后续任务中识别面孔的绩效要显著好于后者 (两组观看视频的总时长相同); 这说明所接触面孔丰富的个人变异性 (within-person variability) 也是影响面孔识别的重要因素” (请见第 1 节第 3 段的蓝字)。

意见 7: “气泡”技术的英文是什么, 怎么做的? 是本文用到的局部遮挡技术吗?

回应: “气泡”技术的英文是“Bubbles task”。根据引用文献的描述, 制作“气泡”刺激的步骤如

下：（1）使用 Matlab 金字塔工具箱（pyramid toolbox）中的拉普拉斯金字塔转换（Laplacian pyramid transform）将一张面孔图片分解成 5 个不同的空间频率波段（106-53, 53-26, 26-13, 13-6, 和 6-3, 剩下的低频波段作为一个恒定背景）。（2）使用不同标准差的高斯孔径（即气泡）在不同空间频率带上进行独立地随机采样，气泡大小根据频段进行调整，气泡数量根据气泡大小调整，最终保持一个给定像素在不同频段带上被揭示的概率不变。这确保了在每个 SF 波段呈现相同的信息量。（3）在气泡遮挡和过滤后的图像之间进行逐点乘法，以获得每个 SF 波段的一个气泡化的脸。（4）将这 5 个随机取样的图像和恒定背景一起叠加，生成了气泡化的刺激，即在试次中呈现给被试的东西。

“气泡”技术与本文所使用的局部遮挡技术有两个相同点：（1）二者都可以帮助判断被试在面孔识别中主要依赖区域。（2）两种技术基于同样的底层逻辑，即对面孔的呈现信息进行了选择，使一部分信息被遮掩，而剩余部分被凸显。两种技术有 4 个区别之处：（1）“气泡”技术对于面孔信息的选择是随机的，只保留了随机生成的气泡区的信息；而局部遮挡技术对面孔信息的选择是固定的，是研究者事先确定好的。（2）“气泡”技术是通过每一试次中气泡位置和被试反应的最小二乘多元线性回归来揭示被试在任务期间使用的区域；而局部遮挡技术是通过不同遮挡情况下被试的反应差异直接判断不同区域在完成任务时的重要性。（3）“气泡”技术由于没有事先对关注区域的选择，因此需要大量的实验试次作为支撑，例如本文引用的 Royer 等人（2016）的研究中就采用了总共 1800 个试次，比较费时费力；而局部遮挡技术由于事先对关注区域进行了选择，只需在所关注的若干条件下进行实验，因此需要的总体试次较少，较为省时省力。（4）“气泡”技术属于数据导向、探索性研究，通过对气泡位置与被试反应的分析得以确定完成任务时的重要区域；而局部遮挡技术属于理论导向、验证性研究，对关注区域有基于前人实验和理论的事前选择，因此研究的发现也可以提供验证性推论。

意见 8：引言倒数第 2 段。“检验面孔部件加工、整体加工和面孔区域的在熟悉度效应中分别起到哪些作用”中的“面孔区域的”是什么意思？

回应：此句中的“面孔区域的”应该是“面孔区域”，多写了一个“的”字。修改稿已删除（请见第 1 节第 7 段的蓝字）。

意见 9：引言倒数第 1 段。“眼睛区域和嘴巴区域的整体加工和部件加工的绩效可能多重模式、复杂的交互效应”不通顺。

回应：原稿漏掉了“表现出”三个字。修改稿将整句改为“眼睛区域和嘴巴区域的整体加工和部件加工的绩效可能表现出多重交互效应”（请见第 1 节第 8 段的蓝字）。

意见 10：本研究结果中某个自变量的两个水平之间的区域外周面孔识别有差异，就得出该自变量改变了被试的整体加工“能力”，似乎不妥。能力一般常用来指比较稳定的表现。

回应：感谢指出。修改稿“整体加工能力”都改为了“整体加工绩效”。

意见 11：“Klatzky 等人在新旧任务中发现，学习阶段既有熟悉面孔又有陌生面孔时，测试阶段中熟悉面孔的识别正确率显著高于陌生面孔(Klatzky & Forrest, 1984)”改成“Klatzky 和 Forrest (1984)在新旧任务中发现，学习阶段既有熟悉面孔又有陌生面孔时，测试阶段中熟悉面孔的识别正确率显著高于陌生面孔”更节省空间节省墨。

回应：感谢建议。修改稿将此句和全文大部分类似句子都改为了引文前置的形式。

第二轮

审稿人 1 意见：

意见 1：很多研究都发现面孔识别中眼睛起了关键作用，本文的相关分析也支持这个结论。但是在对比眼睛和嘴巴间的差异时的一些结果却说明嘴巴更重要（比如注视眼睛时的反应时（ $M = 1307\text{ms}$ ）长于注视嘴巴时（ $M = 1273\text{ms}$ ），表明注视眼睛区域时需要花费更多时间才能正确识别出面孔身份），作者可以解释和讨论下。

回应：我们在修改稿的讨论部分增加了一些讨论（**第四个发现的红字部分**）。具体说来，我们认为这种差异可能是眼睛区域在面孔识别中具有特异性或是锚定作用的反映，因为（1）眼睛重要，所以人们更加擅长在面孔上寻找眼睛而非嘴巴；（2）在注视嘴巴区域时，被试能够快速找到眼睛区域进而形成面孔的知觉表征，所以只需花费较少的时间完成面孔识别（ 1273ms ）；而在注视眼睛区域时，被试需要更多的时间找到嘴巴区域进而形成面孔的知觉表征，所以只需花费较多的时间完成面孔识别（ 1307ms ）。

意见 2：“上述结果表明，人际接触时间对眼睛和嘴巴的识别产生的作用有区域选择性，眼睛区域识别的正确率随接触时间增加而上升，嘴巴区域识别的正确率随接触时间增加而先升后降。”这句话我觉得不妥，作者并没有直接对比接触 1 个月和 13 月组的差异。另外，我也很难理解为什么能得出嘴巴先升后降的结论，希望作者解释下。

回应：感谢指出，这句结论确实存在不妥。得出“嘴巴先升后降的结论”是根据接触 1 个月的被试组在识别熟悉面孔单独嘴巴区域时的正确率显著高于识别陌生面孔单独嘴巴区域，表明 1 个月的接触时间让被试对单独嘴巴区域的识别正确率得到了提升；而接触 13 个月的被试组识别熟悉面孔单独嘴巴区域的正确率与识别陌生面孔单独嘴巴区域相近没有显著差异，表明 13 个月的接触时间反而没有让被试对单独嘴巴区域的识别正确率得到提升。因此我们猜测了单独嘴巴区域识别正确率随接触时间变化的可能趋势：先升后降。但是，这句结论是根据前面分析中两组不同接触时间的被试组在识别眼睛或嘴巴区域时的熟悉度效应变化得出的，属于分析性的比较，并没有直接比较，不够严谨。现在我们补充了对接触 1 个月和 13 个月的被试组从陌生到熟悉区域识别正确率提升差异的直接比较，并对结论中的表达进行了修改（**详见区域部件识别正确率分析最后一段红字部分**）。

意见 3：作者根据相关分析得出的很多结论是基于一个相关显著而另一个相关不显著，而不是两个相关系数间的差异。作者下结论时需要谨慎或是在局限性中提及（作者可以参考这篇论文 *Ten common statistical mistakes to watch out for when writing or reviewing a manuscript*）。此外，对比相关间差异时，没具体提及方法，希望这篇文献能对作者有帮助：[cocor: A comprehensive solution for the statistical comparison of correlations](#)

回应：感谢指出以及提供的帮助，已在相关分析部分和局限性部分做了添加和修改（**见相关分析和讨论最后一段的红字部分**）。

意见 4：第 15 页 11 行缺少一个句号

回应：感谢指出，现已添加“。”号。

意见 5：其他比相关均不显著。多了一个“比”？

回应：感谢指出，“比”字多余，应该为“其他相关均不显著。”

意见 6：“【操作性定义、因变量、被试分配】，”为什么使用“【】”

回应：“【操作性定义、因变量、被试分配】，”为修改时的批注，没有及时删除。给您带来困扰致以歉意，现已删去这些批注。

.....

审稿人 2 意见：

意见 1：正文中的描述和表格中的数据不匹配。请核查是正文有误还是表格有误。如，第 27 页第 6-8 行写道“在识别陌生面孔时，对眼睛区域外周识别正确率显著高于嘴巴区域；在识别熟悉面孔时，对眼睛区域和嘴巴区域外周识别正确率无显著差异。”然而表 5 的数据却不是这样的。第 28 页第 10 行写道熟悉面孔的反应时短于陌生面孔的，然而表 6 中每一个对应的熟悉面孔的反应时都要大于对应的陌生面孔的反应时。

回应：感谢指出。两处均为表格有误，现已核对修改（请见表 5 和表 6 的红字部分）。并将“在识别陌生面孔时，对眼睛区域外周识别正确率显著高于嘴巴区域；在识别熟悉面孔时，对眼睛区域和嘴巴区域外周识别正确率无显著差异。”所在的简单效应分析段落移动到了图 9 和图 10 之间。图 10 是对该结果的直接反映。

意见 2：第 27 页第 9 行总结道：“对嘴巴区域整体加工绩效的提升更明显”，支持这一结论的结果应该是，眼睛上熟悉和陌生的差异，小于嘴巴上熟悉和陌生的差异。然而，作者报告的支持结果却是识别陌生面孔时眼睛和嘴巴有显著差异，识别熟悉面孔时眼睛和嘴巴没有显著差异。

回应：感谢指出，这里我们犯了比较方向上的错误。已在该段结论前添加了从陌生到熟悉后，眼睛区域和嘴巴区域的外周识别正确率提升差异的比较。结果支持了“对嘴巴区域整体加工绩效的提升更明显”这一结论（请见区域外周识别任务正确率分析最后一段）。

意见 3：第 29 页第 12 行“其他比相关均不显著”是什么意思？

回应：“比”字多余，应该为“其他相关均不显著。”

意见 4：参考文献应该首行顶格，其他行缩进，中间不空行。有几个期刊的名称每个字母都是大写的，应每个单词首字母大写即可。

回应：感谢指出，参考文献部分已按照您的意见进行了修改。

第三轮

审稿人 1 意见：

意见 1：“进一步比较从陌生到熟悉，两组被试对单独眼睛区域和嘴巴区域识别正确率提升的差异，发现两组被试在单独眼睛区域识别正确率提升上差异不显著($F(1,55) = 2.494, p = 0.120$)，但在单独嘴巴区域识别正确率提升上差异显著($F(1,55) = 5.161, p = 0.027, \eta^2 = 0.086$)，接触 1 个月的被试组提升($M = 7.2\%$)显著大于接触 13 个月的被试组($M = 0.4\%$)。上述结果提示，人际接触时间对眼睛区域和嘴巴区域的部件识别产生的作用可能具有区域选择性，两区域部件识别正确率随接触时间变化的趋势可能不同：眼睛区域识别的正确率随接触时间增加而上升，嘴巴区域识别的正确率可能随接触时间增加而先上升（接触 1 个月）后下

降（接触 13 个月）。”如果两组被试在单独眼睛区域识别正确率提升上差异不显著，如何下结论说眼睛区域识别的正确率随接触时间增加而上升。

回应：感谢指出。我们在这里的文字和分段不准确，导致了您的困扰。实际上，因为接触时间的增加包括“从陌生面孔与熟悉 1 个月的差异”和“从熟悉 1 个月到熟悉 13 个月”的差异，所以“眼睛区域识别的正确率会随着接触时间增加而上升”是对整段三阶交互效应简单效应分析的一个总结。现稿把这一段重新细写并分为两小段，见下文（**区域部件识别任务正确率分析最后一段绿底黑字部分**）：

对三阶交互效应做进一步的简单效应分析，结果（图 7）发现：接触 1 个月的被试组识别眼睛区域的熟悉度效应边缘显著($F_{(1,55)} = 3.780, p = 0.057, \eta_p^2 = 0.064$)，具体为识别熟悉面孔眼睛的正确率高于陌生面孔；识别嘴巴区域时具有显著的熟悉度效应($F_{(1,55)} = 12.386, p = 0.001, \eta_p^2 = 0.184$)，具体为识别熟悉面孔嘴巴区域的正确率高于陌生面孔。接触 13 个月的被试组只在识别眼睛区域时有显著的熟悉度效应($F_{(1,55)} = 16.171, p < 0.001, \eta_p^2 = 0.227$)，具体为识别熟悉面孔眼睛区域的正确率高于陌生面孔；但在识别嘴巴区域时不存在熟悉度效应($F_{(1,55)} = 0.043, p = 0.837$)。比较接触 1 个月被试组和接触 13 个月被试组识别区域部件的绩效，发现两组被试在单独眼睛区域识别正确率上差异不显著($F_{(1,55)} = 2.494, p = 0.120$)，但是在单独嘴巴区域识别正确率上有所下降，差异显著($F_{(1,55)} = 5.161, p = 0.027, \eta_p^2 = 0.086$)，具体为接触 1 个月的被试组提升($M = 6.2%$)显著大于接触 13 个月的被试组($M = 0.4%$)。

综合起来，上述结果提示，面孔从陌生到熟悉的过程中，熟悉度对区域部件（单独眼睛或嘴巴）识别的效应具有区域异步性和眼睛区域特异性。具体为眼睛区域和嘴巴区域的部件识别正确率随熟悉度变化的趋势不同：眼睛区域识别的正确率随熟悉度累积而逐渐提高，但嘴巴区域识别的正确率随熟悉度增加而先提高（接触 1 个月）后降回到陌生面孔相同的水准（接触 13 个月）。

意见 2：论文经常存在语言上的小错误，作者需要仔细校对全文 比如，“眼睛区域在面孔识别中存在特性的重要性”，什么是‘特性的重要性’？“因此将来的研究可能需要在整个更广泛的接触范围中增加多个取样点来细致考察面孔熟悉过程中加工绩效与内部机制的变化”，因此之后需要逗号。

回应：我们仔细校对了修改稿全文，对语言上的多处小错误进行了纠正。感谢审稿人的意见。

编委意见：I have gone through the revision and the letter. It looks good. Please go ahead and accept the paper for publication.

主编意见：本论文以接触 1 个月和 13 个月的两组大学生为研究对象，借助延迟匹配任务，对两类大学生识别熟悉和陌生面孔的绩效进行了考察。本论文研究设计合理，数据分析过程科学规范，得到的结果具有较为重要的理论价值。