



王晨 | 译 蒋蓉蓉 | 校 万艳 | 编辑

图: 艾略特·林 (ELLOT LIM)

社交网络可能扭曲调研结果。

产品创新 该听网民的吗？

RETHINKING CROWDSOURCING

2

012年，瑞士软饮料公司Rivella打算推出新口味，在开放创新平台上征集消费者建议，共收到800条反馈。在分析过程中，管理者发现有一个提议很受欢迎：研制一款健康的姜味饮料。但公

司调查后发现，原来只是少数几个用户在疯狂吸引点赞和评论，制造声浪。“那只是一小群消费者互相捧场，大量灌水。”负责Rivella创新管道的希尔夫·布劳恩（Silvan Brauen）说。因此，尽管网上评论热度很高，公司认为姜味饮料会遭到失败，放弃了这个方案。

这就是社交偏差的一个案例。新研究显示，企业进行众包式创新、参考消费者意见时，这种认知偏差是一项重要风险。为研究社交偏差对调研结果的影响，瑞士卢塞恩大学营销学教授莱托·霍夫斯泰特（Reto Hofstetter）分析了18家公司的87个众包项目，时间跨度14个月，使用的Atizo是欧洲主流开放创新平台之一，用户包括宝马、雀巢等。本项研究中，霍夫斯泰特的团队共分析了1917名消费者提出的31114项建议。

Atizo系统的一个核心功能是用户投票。研究中的企业平均收到358项建议，分析评估这些创意需要大量管理资源。为减轻企业的负担，Atizo效仿Twitter和Facebook，设置点赞和评论的功能。用户的点赞和评论很受重视：研究中的所有企业都将投票系统作为第一道筛选程序，用以评估创意并决定奖励哪些用户。

但霍夫斯泰特经过分析发现，投票系统并不像看上去那么有价值。像社交媒体上常见的那样，当某个用户的创意被其他人点赞，他通常也会反过来点赞对方的创意。Atizo也有加好友功能，而研究者发现，用户支持好友创意的概率，比支持陌生人高得多。数据显示，社交偏差对创意得到的点赞和评论数影响很大。但研究者发现，企业并不了解这一点。“我完全看不出企业在降低点赞的重要性。”霍夫斯泰特说：“相反，点赞情况对决定奖励和进一步研究哪些创意起着重要作用。”

为进一步分析消费者投票的指导价值，研究者们在众包项目结束一年多后，对企业管理者进行访谈和调查，了解后续进展。研究者随机列出网络参与者的建议，请管理者对每项创意的实用性和创新价值打分。结果显示，消费者青睐的创意与实际转化为成功产品的创意并无相关性。

为找出这种不匹配的原因，研究者邀请145名外部专家，从可行性、原创性和客户价值角度，对所有众包产生的创意打分，然后将得分与网络参与者的意见进行对比。研究者发现，消费者低估可行性的价值，而青睐中等程度的原创性；企业则看重可行性，且要么选择高度原创性的方案，要么选择较为普通的创意。

研究结论可归纳为：“消费者在网上发表的意见，难以作为创意质量的评判依据。”

这并不是说众包没有用；研究者只是建议，企业不应太看重点赞、评论等消费者意见，而应找到更有效评估创意的方法。霍夫斯泰特和同事引用的

● 转 038 页



希尔凡·布劳恩 (Silvan Brauen) “不能只根据流行意见做决策”

希尔凡·布劳恩是瑞士领先软饮料公司Rivella商务拓展负责人，该公司过去5年中曾借助开放创新平台进行产品开发。布劳恩最近接受《哈佛商业评论》采访，介绍了众包创新模式的利弊。以下是采访摘要。

为什么选择众包创新？

我们不想只依靠技术，随意推出一种新口味，然后盼望消费者喜欢。我们想采取新方式，更重要的是，想从消费者需求出发。

效果如何？

非常好。2012年第一次尝试众包创新，我们收到了超过800项创意，其中有我们自己也能想到的普通点子，也有很疯狂的想法，比如甘草口味、颜色奇异的饮料。我们可能会有盲点，或者因为以前失败的方案思维固化，而众包带来的多样性拓展了我们的想象力。

消费者点赞如何影响创意评估？

我们将其视为定性而非定量数据。得到8个赞的创意不一定比得到7个赞的好，这不是严格的科学。但我们认为点赞显示出情感因素或争议，这是积极信号。如果一个创意在众包平台上没有得到关注、激发讨论，那它可能在市场上也不会受到太多注意。

如何处理众包得到的创意？

我们从800个创意中筛选了20个，然后组织内部讨论、焦点小组和品尝测试。这是一个协作、迭代的过程，80%的工作是在网上参与者提交创意后进行的。你不能只依靠流行意见，而应自主思考哪些方案行得通，同时考虑公司整体战略。我们最后决定推出两种新口味：桃子和大黄。这两者都是众包平台上受欢迎度前10%的创意，但不是排名第一的。新产品很成功，带动Rivella在瑞士国内的市场渗透率提高1/3，从30%升至40%。

现在采用开放创新模式是否不如过去多？

是的，而且我觉得其他公司也是。但众包创新变得如此流行，部分原因在于营销而非创新本身。有段时间，如果你宣传说新产品是与消费者合作推出的，反响会相当积极，提高了成功几率。现在很多公司都这样做，所以这不再是卖点。话虽如此，开放创新仍是很有潜力的价值来源。

接 036 页

此前一项研究显示，由不同参与者提出的相似创意（称为“焦点创意”），对产品开发更有帮助。此外，一家众包平台使用更复杂的评价选项代替简单点赞，从而减少了社交性点赞的情形。研究者同时建议众包平台使用算法等工具，排除社交因素的影响。

Atizo 360° 的 CEO 阿德瑞安·盖博（Adrian Gerber）同样认为，企业应减少对消费者投票的关注，而更多依靠自主制定的标准，从几百个创意中筛选出少数几个可行方案。他认为，随着众包的发展，企业开始慎重选择参与群体，进而利用这些人的相关专业知识来引导创新方向。例如，企业正更多邀请员工和供应商而非消费者参与开放创新。还有些企业寻找专家型消费者，参考他们的专业观点。为开发新的防雪崩装备，瑞士户外服装和装备公司 Mammoth 最近就在 Atizo 上召集有设计或工程经验的登山者。这表明，要找到创新解决方案，企业或许应求助合适的小众，而非征集大众的创意。⑥

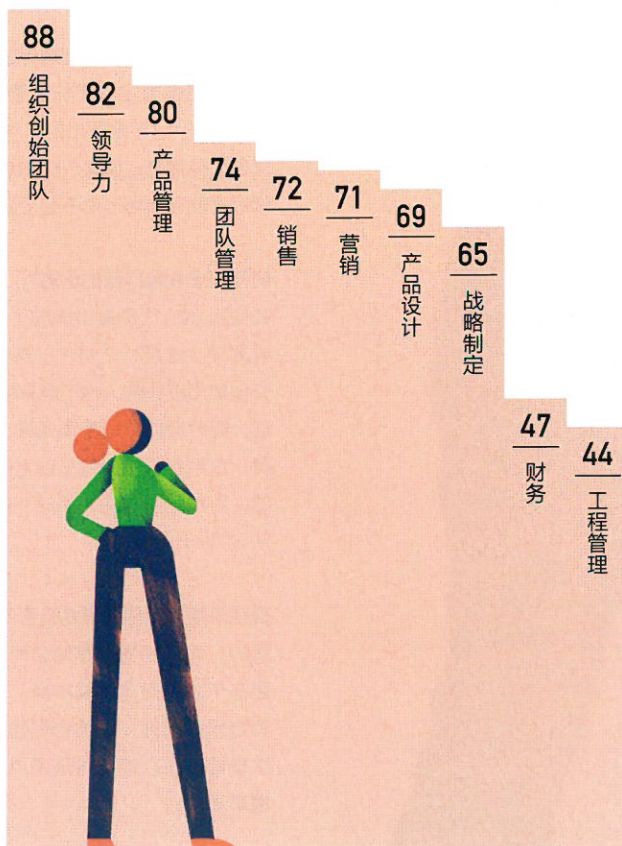
关于本研究 莱托·霍夫斯泰特（Reto Hofstetter）、苏莱曼·阿廖卜塞（Suleiman Aryobsei）、安德里亚斯·赫尔曼（Andreas Herrmann）：《产品创新应该听从网上消费者的意见吗？》（“Should You Really Produce What Consumers Like Online? Empirical Evidence for Reciprocal Voting in Open Innovation Contests”，《产品创新管理》期刊（*Journal of Product Innovation Management*），即将刊登）

创业

创业者应侧重哪些技能？

大学和MBA教育对科技创业有多大帮助？研究者调查了141位曾就读于哈佛商学院的创业者，其中大多数人成立了有VC投资的科技创业公司；此外采访了20位没有MBA学位的创业者。两组受访者都认为，想创业的人应努力成为多面管理者，尤其应培养建立和领导团队的能力，以及发现并响应客户需求的能力；财务和工程等专门技能则被认为重要性较低。一位受访者指出：“这些技能每项都很重要，问题是，CEO自己应当具备哪些能力，而将哪些事情委托给其他团队成员？”■

受访者指出的“重要”或“非常重要”的技能（%）：



资料来源 托马斯·埃森曼（THOMAS R. EISENMANN）、罗伯·豪（ROB HOWE）、贝丝·奥尔特灵格（BETH ALTRINGER）



管理

外向型 CEO 更擅长并购

企业每年花费数十亿美元进行并购，却常常破坏而非创造价值。新研究使用一种新的语言学技巧，分析一项可能影响并购行为和结果的因素：CEO性格外向性。

研究者收集了2381位跨国企业CEO在10年中的季度盈利电话会议上的发言，使用文本分析软件评估他们的性格外向性。之后，研究者分析领导者所在企业的并购表现。结果发现，外向性CEO参与的并购次数更多，标的公司规模更大，且交易的平均回报率更高。

原因何在？相比性格较内向的CEO，外向性CEO参与的董事会更多，因此具备人脉和情报优势，有助于发现和争夺有价值标的。外向性格可能也有助于CEO说服股东认可交易的价值。■

关于本研究 沙文·马尔霍特拉（Shavin Malhotra）等：《性格外向的CEO成功并购概率更大》（“The Acquisitive Nature of Extraverted CEOs”，《管理科学季刊》（*Administrative Science Quarterly*），2017年）



当没有明显答案，或对某个创意的价值看法冲突或犹疑不定时，对照实验可以帮助企业找到方向。

罗恩·科哈维 (Ron Kohavi) 斯特凡·托姆克 (Stefan Thomke) | 文
王晨 | 译 蒋荟蓉 | 校 时青靖 | 编辑

内容

求

打造网站和应用时，很多企
的所有决策，从新产品功能、
观到营销推广，都依靠主观
断而非客观数据。

决方案

业应利用网络对照实验评估
创意。大规模投资有失败风
，而小改变既可能带来不利
响，也可能产生巨大回报，
此企业应对改进方案进行严
测试。

践

理者应学习设计并运行 A/B
试等对照实验，保证其有效
，合理分析结果，并避免错
。

2012 年，微软 Bing 的一名员工想到一个改进搜索引擎展示广告标题的方法。这在技术上并不难，一名工程师几天就可以完成。但同时提出的还有几百个创意，因此，项目管理者并未优先考虑这个改进方案。半年多后，一位工程师看到写代码不太费事，于是发起了一项简单的线上 A/B 测试，检验这个方案的效果。仅仅几小时后，新方案产生了高得不正常的收入，触发了“好过头”警报。通常，这种警报说明可能存在漏洞，但这次不是这样。分析显示，改进方案带来的收入提升高达 12%，相当于美国市场每年增加 1 亿美元收入，而且对用户体验核心指标没有影响。这是 Bing 历史上最赚钱的创意，但直到测试前，它的价值都被低估了。

够丢人的吧！这个例子显示了评估创新方案的潜力是一项多么困难的工作。同样重要的是，它也显示出低成本同时进行多项实验的能力所带来的益处——更多企业已开始认识到这一点。

现在，微软和亚马逊、缤客、Facebook、谷歌等领先企业每年都会进行超过 1 万项网络对照实验，很多测试涉及数百万用户。一些创业公司，以及沃尔玛、赫兹国际租车、新加坡航空等传统企业，也定期进行网络实验，只是规模较小。这些组织发现，“凡事实验”的做法回报惊人。例如，Bing 每个月收获数十项改进创意，每年可将单次搜索产生的收入提高 10% 到 25%；此外，每月还出现数百项能够提升用户满意度的改进方案。这些进步帮助 Bing 保持盈利，且美国个人电脑搜索占有率从 2009 年推出时的 8% 上升至 23%。

如今，互联网对几乎所有企业都非常重要，因此严格的网络实验应被纳入运营流程。如果拥有网络实验所需的软件基础设施和组织能力，企业不仅能评估与网站相关的创意，也能评估商业模式、战略、产品、服务和营销方案，而且成本相对低。对照实验能让决策更加科学、客观，减少直觉成分。如果没有这类实验，很多突破可能永远不会发生，而糟糕创意会得到应用，白白浪费资源。

但我们发现太多组织，包括一些大型互联网企业，在实验方面存在严重缺陷。这些企业不知道如何进行严谨的科学实验，或实验次数过少。

我们研究实验、操作实验，并为多个行业企业提供相关咨询合计已有 35 年。在本文中，我们将分享实验设计和操作、有效性、数据分析等方面的经验，并讨论可能出现的困难。本文将主要探讨最简单的对照实验，即 A/B 测试，但我们的结论和建议也适用于更复杂的实验设计。



认识A/B测试的价值

在 A/B 测试中，实验者对两个对象进行观察：“A”即对照组，一般是当前使用的系统；“B”即实验组，是某种改进方案。如果 A 是“擂主”，那么 B 就是“挑战者”。实验参与者或用户随机体验 A 或 B，二者的核心指标会由计算机进行分析和比较。（与此相对，单变量 A/B/C、A/B/C/D 测试和多变量测试会同时评估多种改进方案。）对于互联网企业，改进方案可能是新增产品功能、用户界面调整（如改进布局）、后台调整（如改进亚马逊图书推荐的算法），或改变商业模式（如提供免运费服务）。无论是销售额、回头率、点击率，还是用户在线时间，企业都可以利用 A/B 测试优化核心运营表现。

只要有几千名日活跃用户，任何企业都可以进行 A/B 测试。如果能采集大量客户样本、自动收集网站和 app 用户的海量互动数据，并同时进行多项实验，企业就能以前所未有的速度和准确度评估大量新创意，而且边际成本接近于零。这使得企业有能力快速迭代、发现错误、完成转变。

认识到 A/B 测试的价值后，一些领先科技企业设立专门团队，负责建立、管理和优化实验基础设施，并在多个产品团队应用。如果使用得当，实验能力可以转化为重要竞争优势。管理者需要留意以下几点：

小改变可能带来大影响。很多人认为，投资越大效果越明显，但网络实验并非如此。更多时候，成功就是把很多小事情做对。虽然商业世界总是将重要的颠覆性创意捧上神坛，但实际上，进步更多是由千百个小改进带来的。

再来看微软的另一个例子。
(本文作者之一科哈维在微软负责实验系统。虽然本文案例大部分来自微软，但能反映很多企业的经验。)

2008 年，微软英国的一位员工提出一项不起眼的建议：用户点击 MSN 主页上的 Hotmail 链接时，默认在新标签页（如果是较老版本的浏览器，则是新窗口）而非原有页面打开邮箱。90 万英国用户参与了新版本测试，结果相当积极：以 MSN 主页点击量计算，用户参与度提升达 8.9%（大多数改进带来的用户参与度提升不到 1%）。然而，对于新版本仍有争议，因为当时很少有网站会在新标签页打开链接，所以新版本只在英国发布。

2010 年 6 月，同样的实验面向 270 万美国用户重复了一次，结果依然良好，因此新版本在全球范围推出。接着，为探索这个创意的其他应用方式，微软尝试将 MSN 搜索结果在新标签页显示。在 1200 万美国用户参与的实验中，用户平均点击数增加了 5%。在新标签页打开链接，是微软实施的提升用户参与度效果最佳的改进，而这一切只需要改几行代码。今天，包括 Facebook 和 Twitter 在内的很多网站都使用这一技巧。

微软的经验并不独特。例如，亚马逊的实验表明，将信用卡产品信息从首页移到购物车页面，每

年可增加数千万美元利润。显然，小投资可能产生大回报。而较大投资却可能回报很低，甚至颗粒无收。微软曾试图将 Bing 与社交媒体整合，在搜索结果页面上增加新窗格，显示来自 Facebook 和 Twitter 的搜索结果。这笔投资高达 2500 万美元，对用户参与度和收入的提升效果却微乎其微。

实验能指导投资决策。管理者可以根据网络测试结果合理分配资金，优化投资决策。例如，微软曾考虑减少 Bing 显示搜索结果所需时间。理论上说，加载速度当然越快越好，但这项改进带来的价值能否量化？应当派 3 名、10 名还是 50 名工程师做这个项目？为回答这些问题，微软进行了一系列 A/B 测试，人为延迟搜索结果加载，并分析速度变化带来的影响。数据显示，加载速度变化 0.1 秒，收入将变动 0.6%。Bing 年收入超过 30 亿美元，因此 0.1 秒的提速能使年收入增加 1800 万美元——足够支持一个较大的团队了。

测试结果还能帮助 Bing 权衡一些重要创新方案，特别是能够提升搜索结果相关度、但降低运行速度的新功能。Bing 不希望很多小功能叠加导致系统性能显著下降。因此，如果某个新功能将导致运行延迟超过几毫秒，微软就会在推出前先提升其性能，或在其他方面提升处理速度。

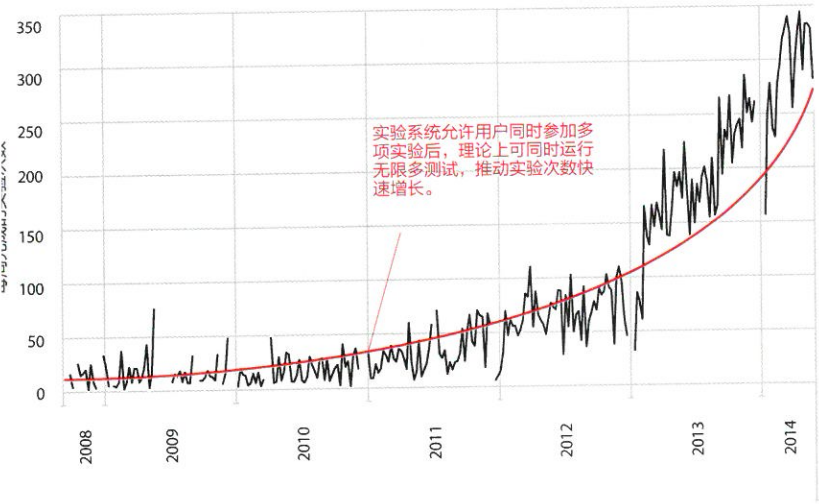
B

构建大规模实验能力

一个多世纪前，经营百货商店的约翰·沃纳梅克（John Wanamaker）说了这句营销名言：“我花在广告上的钱有一半是浪费，但我不知道是哪一半。”我们发现，新创意也差不多是这样：大多数创意都

**亚马逊将信用卡产品信息从
首页移至购物车页面，利润
增加数千万美元。**

3ING实验系统的发展



无法通过实验测试，而即便专家也经常看错有潜力的创意。在谷歌和 Bing，只有 10% 到 20% 的实验会得到积极结果。从微软整体实验情况看，1/3 创意有效，1/3 中性，1/3 产生负面结果。这些都表明，企业须吻很多青蛙（也就是进行大量实验），才能找到王子。

只有凡事实验，才能确保新创意不产生负面影响或意外效果。Bing 内部提出的改进方案中，80% 都要先经过对照实验。（一些低风险的漏洞修复，以及操作系统升级等本地更新不在此列。）

对几乎所有新创意进行科学测试，这需要基础设施，包括插桩技术（记录点击、鼠标悬停、事件时间等）、数据管道，以及数据科学家等。几种便捷的第三方工具和服务能让你做一些尝试，但如果要大规模进行实验，就必须将相关能力嵌入流程，这样才能降低实验的平均成本，并提升可靠性。如果缺少基础设施，实验不仅边际成本较高，也难以取得企业高层的支持。

微软的实验基础设施可称典范——如果企业规模较小或业务不太依赖实验，当然可以降低标准。微软拥有超过 80 人的分析

和实验团队，每天都在进行数百项网络对照试验，涉及产品包括 Bing、Cortana、Exchange、MSN、Office、Skype、Windows、Xbox 等，邀请数十万甚至数百万用户尝试新功能或新变化。团队针对所有测试进行严格统计分析，自动生成的记分卡评估成百上千项指标，并标出重大变化。

企业的实验团队有以下 3 种构成方式：

中心模式。按这种模式，一个数据科学家团队服务整个组织。它的优势在于可以专注长期项目，如打造高质量实验工具、开发先进统计算法等。一个主要劣势是，由于各业务单元优先事项不同，在分享实验团队服务时，可能会在资源分配和成本方面产生冲突。另一个问题是，数据科学家在与业务单元打交道时可能有局外人之感，因此不够认同各部门的目标和专业领域，这让他们难以把握全局、贡献有价值的观点。此外，数据科学家可能对高层缺乏影响力，难以说服他们对必要工具进行投资，或让公司和业务单元领导者认同实验结果。

去中心模式。另一种方法是，将数据科学家分派到各业务单元去。这种模式的好处是，数据科学家可以成为各业务领域的专家。一个主要劣势是，这些专家的职业道路不清晰，而且可能缺少个人发展所需的反馈和指导。此外，各单元独立进行实验，可能无法产生足够有说服力的结果来证明所需工具的必要性。

卓越中心模式。第三个选择，是将部分数据科学家划入中心团队，再把其他人分派到业务单元。（微软使用的是这种模式。）卓越中心主要专注于对照实验的设计、操作和分析。建立全公司范围的实验平台、打造相关工具，可以大大节省实验所需时间和资源。通过组织课程、实验室和会议，卓越中心也能有效传播最佳实验方法。这种模式的一个主要缺陷是，卓越中心和产品团队权责不清晰：当业务单元需要进行更多实验时，谁来出钱请额外的数据科学家？谁又应该负责投资实验结果检验工具？

实验团队的组织方式没有对错之分。小公司一

最优秀的数据科学家遵从特怀曼定律：显得有趣或不寻常的数据通常都是错的。

会先尝试中心模式，或使用第三方工具，规模扩大后再转为其他模式。在业务多元的企业，想要进行重要测试的管理者可能不愿等待总部制定整体规划，这时去中心模式可能更合适，至少一开始如此。如果网络实验属于组织的优先事项，领导者应首先靠中心团队构建相关能力和标准，再将其应用于业务单元。



明确成功标准

企业必须从战略目标出发，定义合适的实验评估指标（一般为一组指标）。这听上去可能简单，但找最能预测长期结果的短期指标实际上很难，很多企业都会出错。为正确建立综合评估标准（overall evaluation criterion, OEC），企业须深思熟虑并进行大量内部讨论。高管层了解战略，数据科学家了解评估指标及其优劣，双方需要密切合作。而且这项工作不应是一次性的——建议每年调整 OEC。

确立 OEC 并不简单。以 Bing 为例，其核心长目标是提升搜索引擎市场占有率和广告收入。有意思的是，降低搜索结果相关度反而会使用户进行多搜索、点击更多广告，从而提升 Bing 的搜索有率和广告收入。但显然，这种收益只是短期的，为用户最终会转而使用其他搜索引擎。那么，哪短期指标能预测长期搜索占有率和收入？经过讨论，Bing 的高层和数据分析师确定了 OEC：尽可能减少用户完成每次任务或会话所需搜索次数，尽可能增加用户的任务或会话次数。

企业还需要将 OEC 的各部分拆解开并进行跟踪，这有助于解释某个创意成功的原因。例如，如

果点击量是 OEC 的核心指标，那么就必须统计页面各个位置的点击量。跟踪不同指标的重要性在于，这能让团队发现实验在其他方面意外产生的影响。例如，对相关搜索项（例如，搜索“哈利波特”，搜索引擎会显示相关书籍、电影、演员阵容等搜索项）进行改动时，由于相关搜索增加，搜索项分布会随之改变，从而对收入产生正面或负面影响。

一段时间后，建立和调整 OEC、分析因果关系会逐渐变得容易。通过进行实验、排查漏洞（下面会讲到）、完成分析，企业不仅会得到哪些指标最适用于哪类实验的宝贵经验，也能设计新指标。Bing 已经设计出超过 6000 项实验指标，并按实验领域归纳成模板（网页搜索、图片搜索、视频搜索、广告改动等）。



留意低质量数据

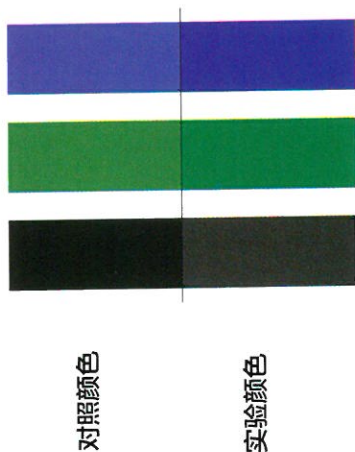
如果实验结果缺少说服力，评估标准再完善也不行。得到数据不难，得到可信任的数据却很难。你需要投入时间和资源来确保实验系统有效，并设置自动检验程序。一种方法是进行严格的 A/A 测试，即自身对照实验，以确保 95% 的时间中不发生显著统计差异。这个简单的方法帮助微软发现了数百项无效实验和不合理的公式使用（例如错误地假设所有统计项彼此独立）。

我们发现，最优秀的数据科学家都具备怀疑精神，并遵从特怀曼（Twyman）定律：显得有趣或不寻常的数据通常是错的。若结果出人意料，实验应当重复一次，这样既确保结果有效，也能消除人们的疑虑。例如，2013 年，Bing 针对搜索结果页

面上不同文本（包括标题、链接、说明文字等）的颜色显示进行了一系列实验。如左图所示，颜色变化非常微小，但结果却意外地积极：标题的蓝色和绿色稍稍加深，说明文字的黑色稍稍变浅，会帮助用户提高搜索成功的几率，且找到相关内容所需时间显著减少。

小改变带来大影响

Bing的实验表明，将标题的蓝色和绿色略微加深，将说明文字的黑色略微变浅，将提升用户体验。新色彩方案全面推出后，年收入可增加1000万美元。



由于颜色变化很难察觉，实验结果自然受到多方质疑，包括设计专家。（多年来，微软像很多企业一样，依靠设计专家而非用户反馈来制定视觉风格、选择色彩。）因此，实验又重复了一次，且参与用户扩大到3200万，得到的结果与之前相似。分析显示，全面推出新色彩方案后，Bing的年收入将增加超过1000万美元。

如果希望实验结果可信，管理者必须保证数据是高质量的。这可能需要排除离群数据、发现采集错误，等等。这点在网络上尤其重要，原因有几点。首先是网络机器人。Bing收到的请求中，超过50%来自网络机器人，这些数据可能扭曲结果或增加“噪音”，令实验者难以判断统计显著性。另一个问题是离群数据点

普遍存在。例如，亚马逊发现，有些个人用户的图书订单太过庞大，导致整个A/B测试结果扭曲，结果是图书馆账户。

如果某些对象群体体现的实验效果明显偏大或偏小（统计学上称为“异质性项目效果”），管理者也应留心。有些情况下，一个异常群组就可能扭曲平均值，使整体实验结果失效。在微软的一次实验中，IE7浏览器用户这一群组由于JavaScript漏洞，无法点击Bing搜索结果，导致整体实验结果由正面变为负面。实验系统应能检测出这类异常群组，

否则实验者只能看到平均结果，错失好创意，也可能无法发现糟糕主意。

如果企业在不同实验中重复使用对照和实验群组，参与者的行为可能受过往经验影响，使结果出现偏差。为避免这种情况，实验者应更换用户群组。

微软实验系统使用的另一项常规检验方法是，验证对照组和实验组用户的实际比例是否符合设计方案。如果二者间出现明显差异，即“样本率不匹配”，经常导致实验结果无效。例如，50.2/49.8（821588名用户对815482名用户）的实际比例与设计方案规定的50/50相差足够大，偶然发生的概率不到1/500000。这种不匹配的情况经常发生（一般每周都有），实验团队需要仔细分析原因并解决问题。



正确认识因果关系

受大数据热潮的影响，一些管理者认为因果关系不重要。他们认为，只需确立相关关系，就可以推出因果关系。错了！

下面两个例子很好地证明了这一点，并揭示出实验缺少对照组的弊端。第一个例子：微软的两个独立实验团队针对Office的两项高级功能进行观察研究，并都得出结论，认为新功能将减少用户流失。实际上，新功能基本都会显示出这种相关性，因为愿意尝试新功能的人大多是重度用户，而重度用户本来就不易流失。因此，新增高级功能可能与低用户流失率相关，却不一定是其原因。收到错误报告的Office用户同样流失率低，因为这些也大多是重度用户。这表明给用户发送错误报告能减少流失

立即下载
《哈佛商业评论》Kindle 版

畅快阅读，轻松管理！

广告



Kindle 版增刊：医改，就该这么改

医改作为全社会普遍关注的话题，自然引发了大量经济、商业和公共政策方面讨论。

Kindle 版增刊：成功避开战略“陷阱”

战略不是运营效率、成本管理和技术创新。运营效率的作用是让企业的一系列运营活动变得更好。

Kindle 版增刊：重思大数据

近年来关于大数据的讨论不绝于耳。所有人都在谈论这种近乎神奇的新资源，而热点也意味着噪音和泡沫，只有能透过迷雾真正把握大数据本质、理解大数据思维的企业才能在这一轮的竞争中赢得先机。

《哈佛商业评论》Kindle 版是《哈佛商业评论》中文版新媒体产品群的重要组成部分。Kindle 版将在每月新上线当期杂志，并新发三本主题增刊。增刊内容将结合商业管理热点，经重新编辑、审校、设计，以最简单的阅读方式提供最优质管理思想，将先进的管理理念传递给中国商界。

电子书定价：3.99 元，
Kindle 电子书优惠价：3.19 元

扫描二维码，
一键下单：



吗？不太可能。

再看第二个例子：雅虎曾进行一项研究，评估雅虎网站上展示的广告能否增加对品牌名称或相关关键词的搜索。实验团队观察到的数据显示，广告使搜索次数增加了 871% 到 1198%。但雅虎又进行了对照实验，发现搜索次数仅增加 5.4%。如果没进行对照实验，雅虎可能认为广告效用巨大，而不会意识到，搜索次数增加实际上来自收集数据期间其他变量的改变。

有些管理者认为，只需证明相关关系，就可以推出因果关系。错了！

显然，观察研究无法确证因果关系。这在医药行业是常识——美国食品药品监督管理局规定，企业必须进行随机临床试验，证明其药品安全有效。

在实验中引入过多变量，不利于理清和分析测试结果，将增大判断因果性的难度。理想情况下，实验应足够简洁，使因果关系清晰可辨。实验设计过于复杂的另一个弊端是，漏洞产生的几率会增加。如果一项新功能导致测试被迫中断的概率是 10%，那么同时引入七项新功能产生致命漏洞的概率就超过 50%。

假如确定 A 是 B 的原因，却不清楚为什么，你应该去探究背后的因果链吗？答案是肯定的。

1500 年到 1800 年，约有 200 万名航海者死于坏血病。今天我们知道，坏血病由膳食中缺乏维生素 C 引起，而水手在长期航行中水果供应不足，容易出现这种状况。1747 年，英国皇家海军的詹姆斯·林德博士（Dr. James Lind）决定加入一次航行，测试 6 种可能的治疗方法。他提供的“解药”包括桔子、柠檬和醋等。结果表明，柑橘类水果能预防坏血病，但没人知道原因。林德误认为水果中的酸是有效成分，因此把果汁加热浓缩以便长期保存，而这恰恰破坏了维生素 C。直到 50 年后，未经加热的柠檬汁纳入每日配给，皇家海军才终于消灭了坏血病。如果当时林德用加热和未加热柠檬汁进行对照试验，

这个发现或许能早些到来，拯救很多生命。

不过还是提醒你，并非一定要知道“为什么”或“如何”，才能从“什么”中受益。当涉及用户行为时尤其如此，因为行为动机可能很难定义。Bing 的一些最重要的突破，其实并没有理论支撑。例如，虽然 Bing 可以通过微调颜色改善用户体验，但并没有成熟理论能解释背后的原因。产生效用的是证据，而非理论。

人们常常认为网络世界动荡凶险，但对照实验可以帮助我们找到方向。当没有明显答案，或对某个创意的价值看法冲突，或犹疑不定时，对照实验能提供指引。

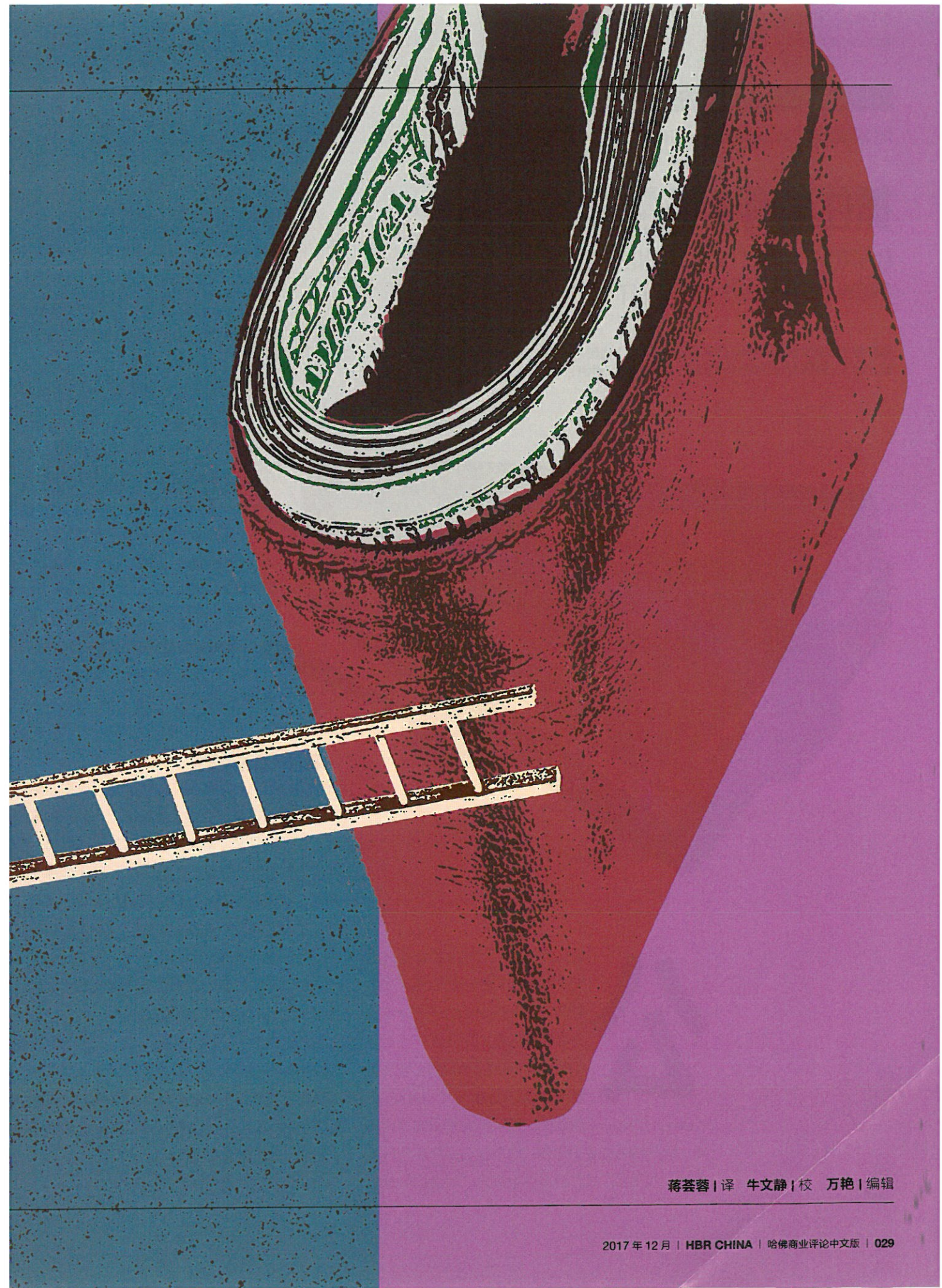
几年前，Bing 曾经讨论是否增大广告展示面积，让广告主增加特定页面的链接。（例如贷款公司可以设置“比较费率”“关于本公司”等链接，而不只是主页链接。）这样做的问题在于，更大的广告会占用更多屏幕空间，导致用户满意度下降、流失增加。团队意见分化，因此进行了一项实验：增大单个广告面积，但减少广告数量，使广告占用屏幕总面积保持不变。结果表明这是一个巨大改进：年收入增加 5000 万美元，而用户体验关键指标未受影响。

如果真想了解一项实验的价值，你可以看一下预期结果与实际结果的差异。如果某事按你的预期发生了，那你并没有学到太多。如果你预期的事情并没有发生，那你学到了重要的东西。而如果你觉得将发生的事情微不足道，但结果却是很大的意外，并且带来突破，那你学到了非常有价值的东西。

将软件的力量与对照实验的严谨结合起来，你的企业可以建立一个学习实验室。在成本、收入、用户体验等方面，它可能给你带来巨大回报。如果想要获得竞争优势，你的企业就应当构建实验能力，并掌握网络实验的科学方法。^⑦



罗恩·科哈维是微软分析与实验团队杰出工程师、总经理。他曾在亚马逊负责数据挖掘和个性化，并负责实验系统 Weblab。**斯特凡·托姆克**是哈佛商学院商业管理 William Barclay Harding 教授。他是商业实验与创新管理领域的权威，与多家全球企业在产品、流程和技术方面进行合作。



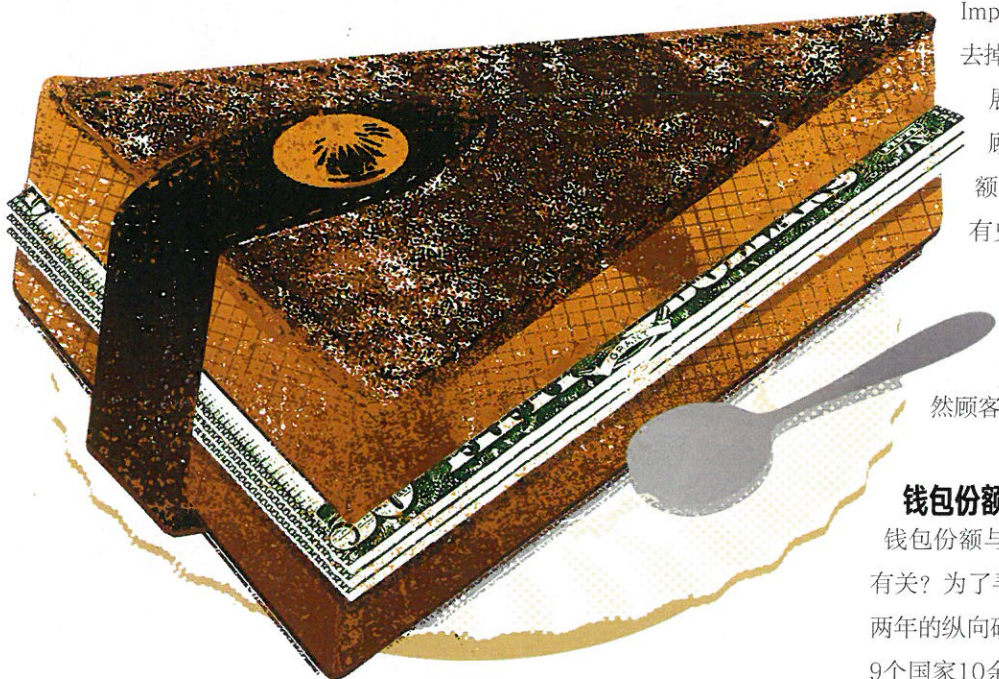
蒋蓉蓉 | 译 牛文静 | 校 万艳 | 编辑

钱包份额法则 (Wallet Allocation Rule) 助你赶超对手。

抓住顾客的钱包份额

CUSTOMER LOYALTY ISN'T ENOUGH.
GROW YOUR SHARE OF WALLET

蒂莫西·凯宁汉 (Timothy L. Keiningham) 勒赞·阿基索 (Lerzan Aksoy)
亚历山大·博耶 (Alexander Buoye) 布鲁斯·库伊 (Bruce Cool) | 文



公

司花费大量时间和金钱，尝试通过评估和掌控客户满意度、净推荐值 (Net Promoter Scores) 等指标提高客户忠诚度。可是，传统的忠诚度评估指标与最重要的“钱包份额” (share of wallet) 之间相关性很低。钱包份额是指客户购买某一类产品时在某一品牌、门店或公司所花费钱的比例。也许客户对你的品牌很满意，愿意向其他人

推荐，但如果这些客户对你的竞争对手同样满意（甚至更喜欢你的对手），你就会损失销量。设法提高顾客满意度，不一定能起作用。这并不是说传统指标没有价值。传统指标可以很好地帮助你了解客户是否满意，是否会向朋友和同事推荐你的品牌，但无法告诉你客户在你的品牌和竞争对手之间如何分配消费。

沃尔玛曾经在这方面受过教训。2008年，沃尔玛根据大量顾客反馈发起了“影响力项目” (Project Impact)，专为改善顾客体验而改造店面，去掉了过道里难看的货盘和货架两端烦人的展示，货架也不再塞得满满当当。结果，顾客满意度果然得到了提升，门店销售额却纷纷下跌，下跌持续时间打破了公司有史以来的纪录。沃尔玛首席财务官查尔斯·霍利 (Charles Holley) 说：“多数顾客还是会来我们店里买东西，但也开始更多地去别的店里买东西。”虽然顾客满意度提升了，钱包份额却有所下降。

钱包份额的重要性

钱包份额与传统的忠诚度指标无关，究竟跟什么有关？为了寻找与之相关的指标，我们开展了为期两年的纵向研究，调查了超过1.7万名消费者，涵盖9个国家10余个行业。我们提出一系列各种各样的问题，收集消费者在一段时间内的购买记录和满意度、忠诚度评价。据我们所知，这是同类研究中规模最大、最严谨的一次。

通过分析，我们找到了一种简洁的相关性。消费者会对自己选择的品牌进行排名，我们总结出一个新的公式——“钱包份额公式”，根据排名预测钱包份额。在各个公司、各个行业，品牌的钱包份额公式得分与其钱包份额都一致，平均相关性超过0.9（完全相关为1.0）。更重要的是，钱包份额公式得分的变化与顾客钱包份额变化的相关性高达0.8。满意度和推荐意愿与钱包份额的相关性只有0.1。

运用钱包份额公式

别被数学吓到了。计算公司的钱包份额，只需要3个步骤，运用一个简单的公式。

1 确定顾客在目标品类中选用的品牌（或商店、公司）数量，比如斯图亚特、玛丽和乔购买洗涤剂都会在Acme、Mega和Brand X这3个品牌里做出选择。

2 调查每个品牌的顾客满意度或其他忠诚度指标得分，按照分数进行排名，出现平局时取平均值，比如有两个品牌并列第一，就以两者都是1.5计。

下表是斯图亚特、玛丽和乔3人对3个洗涤剂品牌的满意度排名。

	Acme	Mega	Brand X
斯图亚特	3	1	2
玛丽	3	2	1
乔	3	1	2

3 计算某位顾客对某个品牌的钱包份额，要将该品牌的排名和该品类的品牌数量代入钱包份额公式：

$$\text{钱包份额} = \left(1 - \frac{\text{排名}}{\text{品牌数量} + 1}\right) \times \left(\frac{2}{\text{品牌数量}}\right)$$

斯图亚特对 Acme 洗涤剂的钱包份额为：

$$\begin{aligned} & \left(1 - \frac{3}{3+1}\right) \times \frac{2}{3} \\ &= (1-0.75) \times 0.67 \\ &= 0.25 \times 0.67 \\ &= 0.1675, \text{约等于 } 17\% \end{aligned}$$

将每位客户和每个品牌的数据代入以上公式计算，便可得到所需数据。要计算某品牌的总体钱包份额，将所有顾客的钱包份额分数相加取平均值即可。

	Acme	Mega	Brand X
斯图亚特	17%	50%	33%
玛丽	17%	33%	50%
乔	17%	50%	33%
品牌钱包份额	17%	44%	39%

钱包份额公式的本质特征是，既考虑品牌在顾客心中的排名，也考虑了顾客在某一品类中选择的牌数量。知道了这两个数值，你就可以稳妥地预钱包份额（计算步骤见附表“运用钱包份额公式”）。

假如消费者考虑某类产品时只有两个备选项，的品牌是其中之一，那么从公式中可以看出，你品牌是首选项还是次选项，会对销售额产生巨大响。在这种情况下，即使你和对手排名并列，后也很严重：你原本可以从顾客那里赚到的钱，有半被竞争对手赚走了。如果顾客的备选项多一些，为次选项的负面影响就没有这么大了。

意义

钱包份额法则可以对战略产生重大影响。要了解钱包份额变化的原因，管理者必须将注意力从影响满度的因素转移到影响排名的因素。

首先，不能在真空中评估品牌表现。这一点似不必再强调，但现实中许多管理者都是这样，评顾客满意度或其他指标时，只参考顾客对本品牌接受度，不考虑其他品牌的存在。这样得出的用考评和奖励管理者的顾客忠诚度目标，只跟顾客意度排名挂钩，无法提升品牌排名，而后者才是多提升钱包份额的因素。

其次，钱包份额法则让我们得以制定直接影响牌表现的战略，评估战略对钱包份额的影响。公用于提升钱包份额的方式往往千篇一律：开展讨顾客的项目，然后评估顾客满意度。看看沃尔玛效训，让顾客更开心，未必能给收入带来积极影响。司应当去了解客户选择各个品牌的原因。如果你

不能在真空中评估品牌表现——然而许多管理者正是这么做的。

的品牌不是首选项，就应该去问顾客，为什么他们更为青睐竞争对手品牌，根据反馈提升自己品牌的排名。钱包份额法则说明，如果无法提升排名，钱包份额就不会提升（见边栏“如何提升排名”）。

下面来看我们研究中的一个综合案例，展示某家提供全面服务的日用百货零售商如何运用钱包份额法则。该零售商开展顾客调查，发现顾客总体对购物体验感到满意，53%的人给“会将这家店推荐给他人”的选项打了9分或10分（满分为10分）的高分。可是，尽管在这方面得分很高，将这家店作为购物首选项的顾客却只有43%。这个结果意味着，57%的顾客更青睐竞争对手，或是认为这家店铺与竞争对手不相上下。

该零售商运用钱包份额法则，计算出本公司和3个主要竞争对手的平均钱包份额，与顾客平均每月杂货开支和选择竞争对手商店的顾客数目相乘，结果显示，3个最大的竞争对手每月从顾客钱包里赚走4.25亿美元——这些钱有一部分可以靠提升排名赢回来。

如何提升排名

想让品牌的排名提升，就要尽量降低客户选择竞争对手的可能性。下面是一个可以迅速应用的简要方法。

运用钱包份额法则，评估客户选择的每个竞争对手所占的钱包份额。

确定客户选择各竞争对手的比例。

计算客户贡献给各竞争对手的收入。

寻找客户选择竞争对手的主要原因。

选择性价比最高的机会：找到客户选择竞争对手的原因，评估一下针对各个原因赢取客户所需的成本，对比赢得客户后可能获得的经济回报。要考虑各竞争对手有共通的地方，解决一些问题可以获得累积效应。

如果你追求的是增长，就要注意排名，别再关注得分。

再回到顾客调查，管理者发现，顾客愿意向别人推荐的原因是商品和氛围都很好。这个发现不出意料，管理团队努力让这两个方面比对手更出众。竞争对手吸引顾客的又是什么？调查显示，一号竞争对手的吸引力在于“天天低价”；二号对手的优势也是价格，主要是轮流大幅度折扣；三号对手则是位置优越。

管理者立刻意识到，要想让店铺在更多顾客心中成为首选项，不能只是继续强化原本的优势。提供质量更好的产品或者继续改进店铺氛围，只会让原本就把店铺作为首选项的顾客们更加满意，不会改变其他注重低价和便利的客人的想法。

不可能在所有品类与对手展开价格战，于是管理者决定给顾客常常购买的日常必需品降价，让喜欢优质商品和店内氛围的顾客少一些光顾低价竞争对手店铺的理由。降价后的顾客调查发现，将这家店作为购物首选的顾客增加了6%，达到49%，顾客经常前往的店铺平均数目从2.5下降到2。将这些变化代入公式可以得出，该零售商钱包份额提升了7%，相当于从对手那里夺回了6200万美元。

如果不再局限于顾客满意度，转而关注利用提升满意度和其他忠诚度的因素赶超竞争对手，许多公司都可以实现这样的增长。如果你追求的是增长，就要注意排名，别再关注得分。通往成功的道路大同小异。重要的不是得到多少分，而是要超越竞争对手。⑥

⑥ 蒂莫西·凯宁汉是益普索满意度和忠诚度（Ipsos Loyalty）业务执行副总裁、全球首席战略官。勒赞·阿基索是福特汉姆大学市场营销学副教授。亚历山大·博耶是益普索满意度和忠诚度业务副总裁。布鲁斯·库伊是范德堡大学欧文管理学院管理学Dean Samuel B. 与 Evelyn R. Richmond 联席教授。