Harvard Business Review

실험 문화 조성

(Building a Culture of Experimentation)

좋은 도구가 아닌 태도를 완전히 바꾸어야 한다.

글쓴이: Stefan Thomke, 하버드대학 경영대학원 교수

번역: 박 영 민

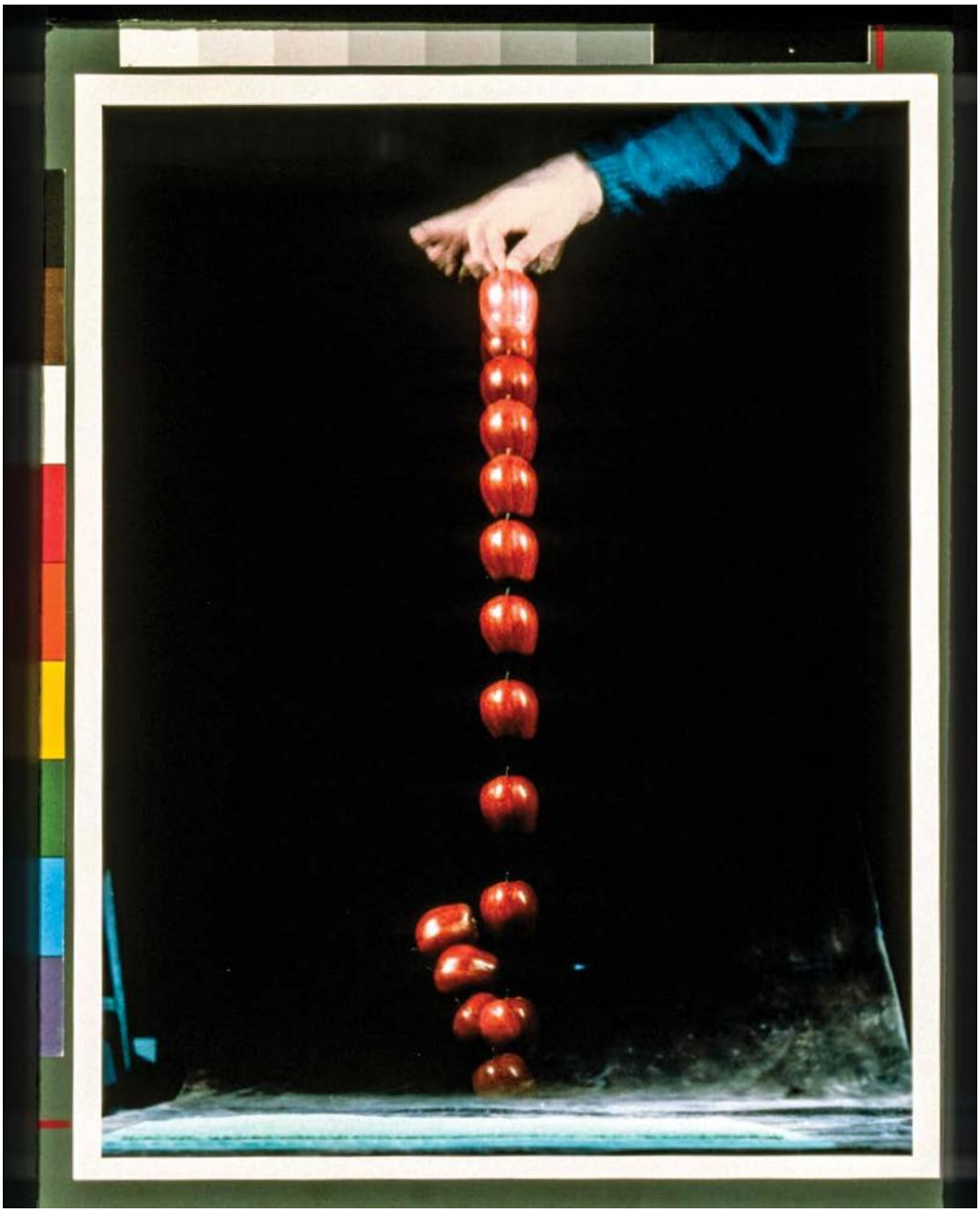
이 글은 <하버드 비즈니스 리뷰(HBR)> 2020 년 3-4 월호에 실린 스테판 톰케 (Stefan Thomke) 하바드경영대학원 교수의 글 “**Building a Culture of Experimentation**” 원문을 번역과 배포에 관한 HBR Publishing Corp 의 Copyright Permission 을 받아 번역하였습니다.

2017년 12월, 바쁜 휴가 여행 시즌 직전, Booking.com의 디자인 책임자는 급진적인 실험을 제안했다. 회사 홈페이지에 완전히 새로운 레이아웃을 테스트하는 것이었다.

기존 홈페이지에서처럼 호텔, 휴가 용 숙박 시설, 여행 상품에 대한 다양한 옵션을 제공하는 대신, 새 홈페이지는 고객이 어디로 가는지, 날짜와 파티에 참석하는 사람 숫자를 묻는 작은 창만 표시하고, "숙박 시설", "항공편", "렌터카"라는 세 가지 간단한 옵션을 제시했다. Booking.com에서 수년간 최적화를 도모해온 콘텐츠와 디자인 (사진, 텍스트, 버튼, 메시지)을 없앴다.

당시 Booking.com 의 CEO 였던 질리언 탠스는 회의적이었다. 그녀는 이 변화가 회사의 충성 고객들에게 혼란을 줄까 걱정했다. 당시 회사의 핵심 실험 팀장이었던 루카스 베르미어는 샴페인 한 병을 걸고 테스트가 "실패" 할 것이라고 내기를 걸었다. 즉, 회사의

중요한 성과 지표인 고객 전환율, 즉 웹사이트 방문자의 예약 비율이 떨어질 것이라는 의미였다. 그런 비관적인 상황에서 왜 최고경영진은 그냥 실험을 거부하지 않았을까? 그렇게 하면 Booking.com 의 핵심 원칙 중 하나인 회사의 모든 사람이 경영진의 허가 없이 무엇이든 테스트할 수 있다는 원칙을 위반하는 것이기 때문이었다.



Booking.com 은 1,000 개 이상의 엄격한 테스트를 동시에 실행하고, 나의 추정에 따르면 1 년에 25,000 개 이상의 테스트를 실행한다. 특정 시점에 수천억 개의 랜딩 페이지의 다양한 버전이 있어서 동일한 장소에 있는 두 고객이 같은 버전을 볼 가능성이 매우 낮다는 것을 의미한다. 이 모든 실험은 20 년도 채 안 되어 작은 네덜란드의 스타트업에서 세계 최대의 온라인 숙박 플랫폼으로 회사를 탈바꿈하는 계기가 되었다.

Booking.com 은 온라인 실험의 힘을 발견한 유일한 회사가 아니다. Amazon, Facebook, Google, Microsoft 와 같은 디지털 거대 기업들은 마케팅과 혁신에 있어서 온라인 실험이 게임 체인저 라는 것을 발견했다. 예를 들어, Microsoft 의 Bing 사업부가 매월 수십 개의 개선을 하였고, 이를 통해 검색 당 수익이 연간 10%~25% 증가했다. (2017 년 9~10 월 HBR 에 실린 "The Surprising Power of Online Experiments" 참조) FedEx, State Farm, H&M 등 디지털 기반이 없는 회사도 온라인 테스트를 도입하여 최상의 디지털 접점, 디자인 선택, 할인, 제품 추천을 파악했다.

“점점 더 디지털화되는 세상에서 장기적으로, 그리고 많은 산업에서 단기적으로 대규모 실험을 하지 않으면 세상에서 사라지게 됩니다." Expedia Group 의 CEO 인 마크 오커스트롬이 내게 말했다. "우리는 항상 수백만 명의 방문자를 포함하여 수백, 아니 수천 개의 동시 실험을 진행하고 있습니다. 이 덕분에 고객이 무엇을 원하는지 추측할 필요가 없습니다. 우리는 존재하는 가장 방대한 '고객 설문 조사'를 반복해서 실행하여 고객이 무엇을 원하는지 말하게 할 수 있는 능력을 갖추고 있습니다.”

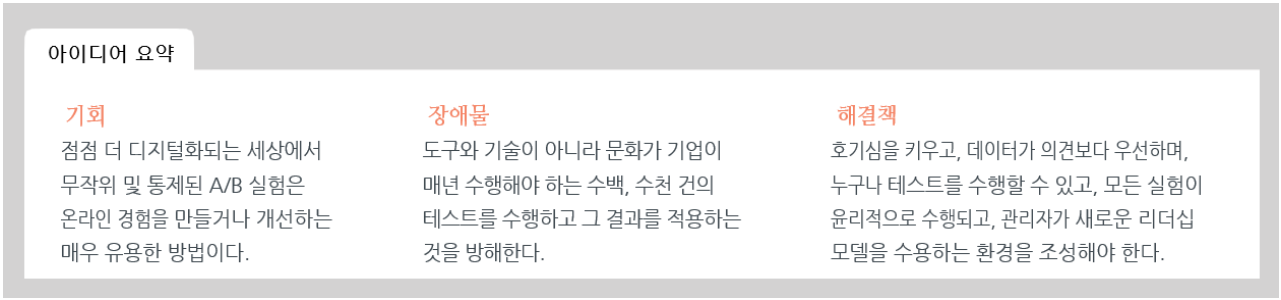
하지만 수십개의 조직을 연구하고 1,000 개가 넘는 실험에 대한 익명화 된 데이터를 분석한 결과, Booking.com, Expedia 및 이와 유사한 조직은 예외라는 것을 알게 되었다. 많은 회사는 1 년에 수백 또는 수천 건의 온라인 테스트를 실행하는 대신 영향이 거의 없는 수십 건을 넘지 않게 운영한다.

테스트가 그렇게 가치 있다면, 왜 기업들은 더 많이 하지 않을까? 이 질문을 수년간 조사한 결과, 핵심적인 이유는 문화라고 이야기하고 싶다. 기업들이 온라인 실험 능력을 확장하려고 시도할 때, 종종 장애물이 도구와 기술이 아니라 공유된 행동, 신념, 가치라는 것을 알게 된다. 성공하는 실험이 하나 있을 때마다 거의 10 개가 실패한다. 효율성, 예측 가능성, "승리"를 강조하는 많은 조직의 눈에는 이러한 실패는 낭비이다.

혁신을 성공적으로 이루려면 기업은 예산이 부족할 때조차도 실험을 일상의 필수적인 부분으로 만들어야 한다. 즉, 직원의 호기심을 키우고, 데이터가 의견보다 우선하며, R&D 에 있는 사람뿐 만 아니라 누구나 테스트를 수행하거나 의뢰할 수 있고, 모든 실험이 윤리적으로 수행되며, 관리자가 새로운 리더십 모델을 수용하는 환경을 조성해야 한다. 이 글에서는 이러한 일을 잘 해낸 몇몇 기업을 살펴보겠다. 특히, 발견된 가장 강력한 실험 문화를 가진 Booking.com 에 초점을 맞추었다.

호기심 키우기

조직의 모든 사람, 경영층부터 실무진 까지, 예산을 할당이 어렵고 언제, 얼마나 자주 발생할지 예측할 수 없지만, 놀라움(Surprises)을 소중히 여겨야 한다. 회사가 이것을 받아들이면 호기심을 갖게 되고 사람들은 실패를 값비싼 실수가 아니라 학습의 기회로 볼 것이다.



전형적인 사례로는 모바일 기기용 게임인 Air Patriots 의 개정판과 관련된 Amazon 의 사건이 있다. 이 게임에서 플레이어는 비행기 편대를 이용해 타워를 공격으로부터 방어한다. Amazon 에서 새 버전을 출시했을 때 개발팀은 반응에 당황했다. 7 일 사용자 유지율이 놀랍게도 70%나 떨어졌고 매출은 30%나 감소했다. 팀은 실수로 게임의 난이도가 약 10% 높아졌다는 사실을 발견했다. Amazon 은 재빨리 수정판을 출시했지만 개발자들은 게임을 더 쉽게 만들면 유지율과 매출이 크게 증가할지 궁금해했다. 이를 확인하기 위해 그들은 통제 외에도 4 가지 새로운 난이도 레벨로 테스트를 실행했고 가장 쉬운 변형이 가장 좋은 성과를 거두었다는 사실을 알게 되었다. 몇 가지 추가 개선을 거친 후 Amazon 은 새 버전을 출시했다.

이번에는 사용자가 20% 더 오래 플레이했고 매출이 20% 증가했다. 사고로 인해 놀라운 통찰력을 얻었고, 이는 새로운 실험의 출발점이 되었다.

불행히도 이런 종류의 반응은 예외이다. 많은 회사에서 실험과 관련된 위험 때문에 관리자들은 그들에게 자원을 할당하는 것을 꺼린다. 하지만 믿음의 도약을 한 회사들이 누리는 이익은 다른 사람들에게 그들을 따를 용기를 줄 것이다.

많은 조직은 또한 실험의 본질과 양에 대해 너무 보수적이다. 성공적인 실험의 중요성을 지나치게 강조하면 직원들이 익숙한 솔루션에 집중하거나 직원들이 이미 작동이 잘 될 것으로 아는 솔루션에 집중하게 하고 실패의 두려움이 존재하는 아이디어는 테스트하지 않도록 유도하게 하는 점이 있다. 그런데 실제로 적은 수의 실험보다 많은 수의 실험을 하는 것이 덜 위험하다.

Booking.com 에서 긍정적인 결과를 도출하는 실험은 약 10%에 불과하다. 긍정적 결과란 판매, 반복 사용, 클릭률 또는 사용자가 사이트에서 소비하는 시간 등을 개선하는 수정안인 "B"에 대한 사용자들(무작위로 할당된)의 반응이 현 상태인 "A"에서의 반응보다 긍정적 결과를 보여 주는 것이다. (A/B 테스트 외에도 Booking.com 은 동시에 두 개 이상의 수정을 평가하는 보다 복잡한 테스트도 실행한다.)

그러나 많은 양의 실험을 수행할 때 낮은 성공률이지만 상당한 수의 성공적 결과물이 나올 수 있으며, 이는 결국 실패로 인한 재정적 정서적 비용을 능가하게 되는 결과를 낳게 한다. 한 회사가 일 년에 몇 번의 실험만 한다면, 단 한 번의 성공만 거둘 수도 있고, 운이 나쁘다면 한 번도 성공하지 못할 수도 있다. 그렇다면 실패는 큰 문제가 된다.

본인이 연구한 회사에서는 완전히 새로운 제품을 개발하는 초기 아이디어의 성공률이 매우 낮았다. 그러나 초기 실패를 통해 개발자는 불리한 옵션을 신속하게 제거하고 더 유망한 대안에 노력을 집중할 수 있었다.

실험적 문화에서 직원은 실패 가능성에 굴하지 않는다. "여기에서 성공하는 사람들은 호기심이 많고, 열린 마음을 가지고 있으며, 배우고 문제를 파악하려는 열망이 강하고, 틀렸다는 것이 증명되는 것을 감수합니다." 현재 Booking.com 에서 모든 테스트를 이끄는 Vermeer 의 말이다. 이 회사의 채용 담당자는 그런 사람을 찾고 있으며, 그들이 본능을 따를

수 있도록 권한을 부여하기 위해 회사는 신입 사원을 엄격한 온보딩 프로세스에 참여시키고, 여기에는 실험 교육이 포함되며, 그런 다음 모든 테스트 도구에 대한 접근 권한을 제공한다.

데이터가 의견보다 우선한다고 주장

온라인 실험의 경험적 결과는 누구의 의견이든 강한 의견과 충돌할 때 존중되어야 한다. 이것이 Booking.com 의 태도이다, 그러나 이해할 만한 여러 이유 때문에 대부분의 회사에서는 그렇지 않으며 이것은 인간의 본성 때문이다. 우리는 편견을 확인하는 "좋은" 결과를 기꺼이 받아들이지만, 우리의 가정에 맞지 않는 "나쁜" 결과는 수긍하지 않고 철저히 조사하는 경향이 있다.

해결책은 예외적인 경우를 빼놓고 실험에서 검증된 사항을 반영하여 구현하는 것이다. Booking.com 의 임원이 말했듯이, "테스트 결과에서 웹사이트 제목이 분홍색이어야 한다는 결과가 나왔으면 분홍색으로 바꾸어야 합니다. 항상 테스트 결과를 따라야 합니다."

최고경영층이 이 규칙을 따르도록 하는 것은 쉽지 않다. (미국 작가 업튼 싱클레어가 한때 "사람의 월급이 어떤 것을 이해하지 않는 데 달려 있다면, 그것을 이해시키기는 어렵다" 라고 농담조로 말했듯이) 하지만 그들이 그렇게 하는 것이 중요하다. 소위 HiPPO(highest-paid person's opinion: 최고 급여를 받는 사람의 의견)만큼 혁신을 더 빨리 멈추게 하는 것은 없다.

모든 의사결정이 온라인 실험에 기반할 수 있거나 기반해야 한다고 말하는 것은 아니다. 어떤 것들은 테스트를 실시하기 매우 어렵거나 불가능하다. 예를 들면, 회사 인수에 관한 전략적 결정과 같은 경우이다. 하지만 온라인에서 테스트할 수 있는 모든 것을 테스트한다면 실험은 의사 결정에 도움이 되고 건강한 토론을 촉진할 수 있다.

때때로 이러한 토론은 데이터를 무시하려는 의식적인 선택으로 이어질 수 있다. 넷플릭스에서 코미디 시리즈와 관련된 한 가지 결정에서 일어난 일이 바로 그것이다. 넷플릭스는 대규모 실험을 위한 정교한 인프라를 구축했다.

(안드레아스 파이닝거(Andreas Feininger)의 장시간 노출 사진은 헬리콥터의 빛의 궤적을 포착한 것임.)

2018 년에 발행된 월스트리트 저널 기사에 따르면, 그레이스와 프랭키의 스타 중 한 명인 릴리 톰린만 등장하는 프로모션이 톰린과 그녀의 공연자 제인 폰다가 등장하는 프로모션보다 잠재 시청자의 클릭 수가 더 많다는 실험 결과가 나왔을 때, 이 회사의 임원진은 갈등을 겪었다. 콘텐츠 팀은 제인 폰다를 제외하면 여배우가 소외되고 계약 위반이 될 수 있다고 우려했다. 경험적 증거와 "전략적 고려 사항"을 놓고 격렬한 논쟁이 벌어진 후, 넷플릭스는 고객 데이터가 결정을 뒷받침하지 않았음에도 불구하고 두 여배우가 포함된 이미지를 사용하기로 했다. 그러나 실험적 증거는 이러한 상쇄 관계(Trade off)를 더 투명하게 만들었다.

실험의 민주화

앞에서 언급했듯이, Booking.com 의 모든 직원은 경영진의 허가 없이 수백만 명의 고객을 대상으로 실험을 시작할 수 있다. 1,800 명의 기술 및 제품 담당 직원 중 약 75%가 회사의 실험 플랫폼을 적극적으로 사용한다. 표준 템플릿을 사용하면 최소한의 노력으로 테스트를 설정할 수 있으며 사용자 모집, 무작위화, 방문자 행동 기록 및 보고와 같은 프로세스는 자동화된다. 핵심 실험 팀과 5 개의 위성 팀은 전체 조직에 교육과 지원을 제공했지만 회사의 요구 사항이 변화함에 따라 해당 구조는 최근 베르미어에게 보고하는 4 개의 중앙 팀과 제품 팀에 배치된 전문가("앰버서더")로 변경되었다.

실험을 진행하기 위해 개인이나 팀은 모든 사람이 볼 수 있는 전자 양식을 작성해야 한다. 이 양식에는 실험 이름, 목적, 주요 수혜자(고객 또는 공급업체), 관련된 과거 실험, A/B, A/B/C 또는 A/B/n 테스트에서 시도할 수정 횟수 등이 포함되어 있다.

실험이 시작되고 실행되면 팀은 처음 몇 시간 동안 실험을 면밀히 관찰한다. 1 차 또는 2 차 지표가 빠르게 떨어지면 팀은 테스트를 중단할 수 있다. 그 초기 시간 이후에도 플랫폼은 계속해서 자동으로 데이터 품질 검사를 실행하고 무언가 이상하면 경고 메시지를 보낸다. 개방성을 장려하기 위해 Booking.com 은 성공, 실패, 반복 및 최종 결정에 대한 전체 설명이 포함된 과거 실험의 중앙 검색 가능 저장소를 유지 관리한다. 그리고 모든 사람이 진행 중인 실험에서 생성된 실시간 정보를 볼 수 있다.

"다소 아이러니하게도, 실험 인프라를 중앙 집중화 하는 것이 조직적 분산화를 가능하게 합니다." 베르미어가 내게 설명했다. "모두가 같은 도구를 사용합니다. 이를 통해 서로의 데이터에 대한 신뢰가 높아지고 토론과 책임성이 가능해집니다. Microsoft, Facebook, Google 과 같은 일부 회사는 머신 러닝과 같은 분야에서 기술적으로 더 발전했을 수 있지만, 간단한 A/B 테스트를 사용하면 모든 사람을 참여시키는 데 더 성공적입니다. 조직 전체에서 테스트를 민주화 했습니다."

물론 민주화에는 어려움이 따른다. 그중 하나는 팀이나 개인이 Booking.com 의 트래픽이 많은 웹사이트에서 무언가를 망가뜨려 다운 될 수 있는 위험이다. 또 다른 하나는 각 팀이 자체 방향을 설정하고 어떤 사용자 문제를 해결하고 싶은 지 파악해야 한다는 것이다. 이를 위해서는 광범위한 교육과 올바른 문제가 무엇인지에 대한 팀원 간의 지속적인 토론이 필요하다. 토론이 장려되고, 사람들은 의심스러운 것이 있으면 동료에게 연락한다. 누구나 실험을 시작할 수 있는 것처럼 누구나 실험을 중단할 수 있다.

그러나 이는 실험이 치명적으로 잘못되었을 때에만 드물게 발생한다. 예를 들어, 누군가가 밤에 사무실에 혼자 있고 실험 중에 고객 전환율과 같은 주요 지표가 급락한 것을 발견한 경우이다. (실험을 계속하면 회사에 수백만 달러의 손실이 발생)

이 시스템은 팀에 가치 있다고 믿는 새로운 접근 방식을 시도하는 데 필요한 자율성을 제공하고 회사 전체의 사람들이 실시간으로 실험을 모니터링하고 피드백을 제공할 수 있도록 한다. 모든 사람이 Booking.com 의 비즈니스를 개선하는 방법에 대한 아이디어를 테스트할 수 있도록 진정으로 해방시켜준다.

윤리적으로 민감 하라

새로운 실험을 고려할 때 기업은 사용자가 다음을 고려할 것인지 여부에 대해 신중하게 생각해야 한다. 답이 항상 명확한 것은 아니지만, 이 질문을 검토하지 않는 조직은 반발을 불러일으킬 위험이 있다. 2012 년 페이스북은 일주일 동안 진행한 실험을 통해 감정 상태가 플랫폼에서 전염되는지 여부를 알아봤다. 페이스북은 알고리즘에 따라 큐레이팅된 게시물, 스토리, 활동 목록인 뉴스 피드를 재구성하여 긍정적인 뉴스 스토리를 덜 보면 사람들이 긍정적인 게시물의 수를 줄이는지 확인했다.

그 네트워크는 또한 사람들이 부정적인 뉴스에 덜 노출되었을 때 그 반대가 일어났는지도 테스트했다. 이 실험에는 무작위로 선택된 약 690,000 명의 사용자가 참여했으며, 그 중 약 310,000 명은 자신도 모르게 뉴스 피드에서 조작된 감정 표현 게시물을 접하게 하고 나머지는 접하지 않게 하였다.

페이스북과 코넬 대학의 연구진이 학술지에 연구 결과를 발표하자 대중의 분노가 폭발했다. 페이스북의 데이터 사이언스 팀은 수년 동안 논란 없이 의심을 받지 않는 사용자들을 대상으로 실험을 해왔지만, 그 감정적 조작은 신경을 건드렸다. 비평가들은 페이스북의 일반적인 데이터 사용 정책에 대한 참가자들의 동의가 충분한지에 대한 우려를 제기했다. 그들은 회사가 사용자가 테스트를 거부할 수 있다는 점을 명확하게 하고 연구를 위해 데이터를 수집한다는 것을 보다 더 명확하게 했어야 한다고 생각했다. 학습의 관점에서 볼 때, 이 실험은 성공적이었다: 비록 그 영향은 매우 작았지만, 온라인에 감정적 전염이 존재한다는 것을 발견했다. 그러나 일부 사용자는 페이스북이 과학이라는 명분으로 자신을 이용했다고 느꼈다.

연구에 따르면 새로운 아이디어를 먼저 테스트하는 기업은 실험을 수행하지 않고 새로운 프랙티스를 구현하는 경쟁업체보다 더 많은 고객 감시에 직면한다. 의료, 자동차 디자인, 세계 빈곤과 같은 분야의 16 개 연구를 공개적으로 분석한 결과, 생명윤리학자 미셸 마이어와 그녀의 동료들은 참가자들이 A/B 테스트가 테스트 없이 바로 전체를 대상으로 실행에 옮기는 프랙티스 보다 도덕적으로 더 의심스럽다고 생각한다고 결론지었다. 두 가지 치료법이 모두 반대할 수 없는 경우에도 마찬가지였다.

분명히 윤리 교육과 모종의 감독이 필요하다. 사람들이 지나치게 조심하지 않도록 하거나 형식적인 레드 테이프에 얽매이지 않는 방식으로 감독을 수행하는 것은 도전적인 일이다. 이러한 정확한 이유 때문에, Booking.com 은 어떤 종류의 테스트가 가능한지에 대해 상부에서 규칙을 부과하는 것을 주저해 왔다. 대신, 직원들이 실험이나 제안된 관행이 고객에게 도움이 될지 해가 될지 물어보도록 권장한다.

Booking.com 의 최고 제품 책임자인 데이비드 비스만스는 "차라리 감시나 윤리검토 위원회에서 벗어나고 싶습니다."라고 말했다. "그것은 확장 가능한 솔루션이 아닙니다. 병목 현상이 생길 수 있고, 테스트 감시를 한다고 해서 사람들이 힘을 얻었다고 느끼지 못합니다"고 말했다. 대신 회사는 모든 직원에게 열려 있는 내부 온라인 포럼에서 토론을 권장한다. 논쟁은 활기 있게 전개될 수 있으며 고객이 거래를 마치도록 설득하는 기술을 사용하는 것과 같은 문제를 다룰 수 있다. (예: "지금 예약하지 않으면 이 기회를 잃게 됩니다 " 또는 "방 세 개만 남았습니다" 와 같은 메시지). "나는 차라리 스스로 교정하는 공동체가 낫다고 생각한다." 라고 비스만스는 설명했다.

이를 위해 Booking.com 의 온보딩 프로세스에는 윤리 교육도 포함된다. 대규모 실험 프로그램을 운영하는 또 다른 회사인 LinkedIn 은 약간 다른 접근 방식을 취한다. 회사는 "부정적인 회원 경험을 제공하거나, 회원의 기분이나 감정을 바꾸려는 목표를 갖거나, 기존 회원의 설정이나 선택을 무시하려는 목적"의 실험을 실행하지 않을 것이라고 명시한 내부 지침을 만들었다.

다른 리더십 모델 도입

실험을 민주화하고 테스트 결과를 따름으로써 회사는 직원이 스스로 좋은 결정을 내리게 하고 혁신과 개선을 가속화 하게 된다. 하지만 대부분의 결정이 이런 방식으로 내려진다면, 회사의 전략적 방향을 개발하고 회사 인수와 같은 큰 결정 외에 상위 리더가 할 수 있는 일이 무엇이 있을까? 최소한 다음 네 가지가 있다.

검증 가능한 가설과 핵심 성과 지표로 나눌 수 있는 큰 도전 과제를 설정하라.

직원들은 실험이 전반적인 전략적 목표를 어떻게 지원하는지 알아야 한다. Booking.com 의 고위 임원들이 직원들에게 업계 최고의 온라인 경험을 디자인하도록 도전했다고 가정해 보겠다. 그들은 더 나은 경험이 더 많은 고객 트래픽을 생성하여 더 많은 공급업체를 Booking.com 플랫폼으로 유치하고 고객 기반과 활동을 더욱 확대하는 데 도움이 될 것이라고 기대할 수 있다. 그 목표를 달성하는 방법을 발견하기 위해 직원들은 가설과 관련 지표를 고안할 수 있다. 예를 들어, 중요한 텍스트에 밑줄을 긋는 것은 중요한 정보를 찾기 쉽게 만들어 전환율을 높이고 "한 번의 클릭, 무료" 취소 옵션은 순 호텔 예약이 감소하지 않으면서 사용자 복귀율을 높일 것이다.

시스템, 자원, 대규모 실험을 허용하는 조직 설계.

거의 모든 아이디어를 과학적으로 테스트하려면 계측, 데이터 파이프라인 및 데이터 사이언티스트와 같은 인프라가 필요하다. 다양한 외부 도구 및 서비스를 활용하면 실험을 쉽게 시도할 수 있지만, 규모를 확장하기 위해 고위 경영진은 테스트 기능을 회사 프로세스에 긴밀하게 통합해야 한다. 이를 위해서는 집중화(centralization)와 분산화 (decentralization) 사이에서 적절한 균형을 유지해야 한다.

집중화 그룹에서는 개발자, 사용자 인터페이스 디자이너 및 데이터 분석가와 같은 전담 전문가가 회사 전체를 대상으로 실험을 진행하고 최첨단 방법 및 도구를 도입하는 데 집중할 수 있다. 그러나 테스트가 소수의 전문가 그룹에 국한된다면 실험을 확장하기 어려울 것이다. 분산형 테스트에서 기업은 여러 사업부에 걸쳐 전문가 팀을 분산시킨다. 이 접근 방식은 실험을 조직의 더 많은 부분으로 확장하지만 지식 공유를 방해하고 상충되는 목표와 전문가 간의 조정 부족으로 이어질 수 있다.

처음에는 더 넓은 조직을 참여시키기 위해 탈집중화가 필요할 수 있지만, 그 후에는 기업이 실험 능력을 향상시키는 데 눈을 돌려야 한다. 그것이 Booking.com 이 한 일이다. 처음에는

위성 팀을 통해 회사 전체에 실험을 분산시켰지만, 사용자를 지원하느라 너무 바빠서 전사적인 역량을 구축하는 데 집중할 수 없다는 것을 알게 되었다. 이 문제를 해결하고 팀을 더 잘 조정하기 위해, Booking.com 최근 사업부를 지원하고, 실험에 대한 회사의 접근 방식을 표준화하고, 모범 사례를 채택하고 따르도록 하는 CoE(Center-of-Excellence) 모델로 전환했다.

롤 모델이 되라

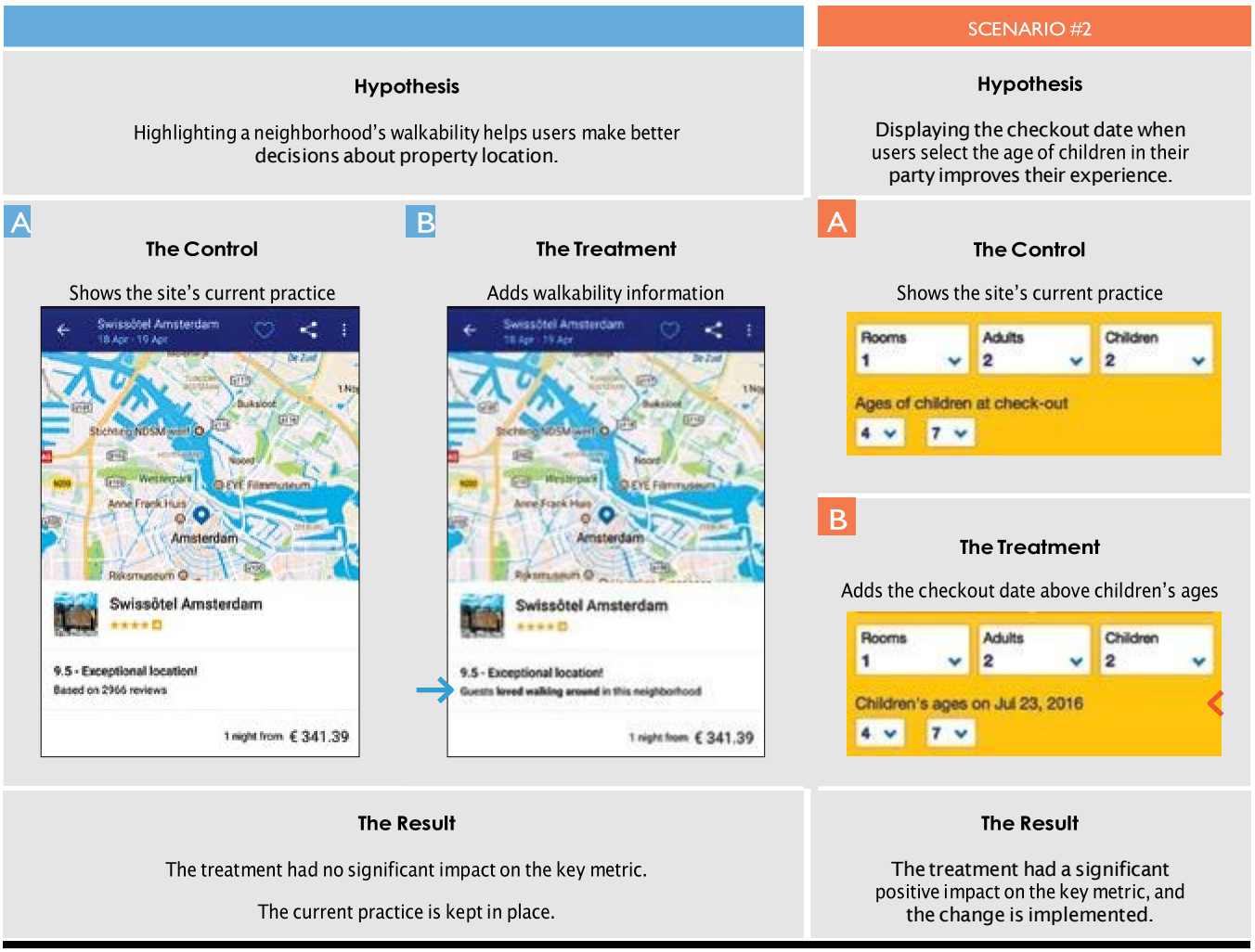
리더는 다른 모든 사람과 같은 규칙에 따라야 하고 자신의 아이디어를 테스트에 적용해야 한다. "자존심을 갖고 항상 자신이 가장 잘 안다고 생각해서는 안 됩니다." 텐스가 나에게 말했다. "제가 CEO 로서 누군가에게 '우리 사업에 도움이 될 것 같아서 이렇게 하라고 합니다' 라고 말하면 직원들은 문자 그대로 저를 보고 '좋아요, 좋습니다. 테스트해서 당신이 옳은 지 알아보겠습니다' 라고 말할 것입니다." 상사는 지적인 겸손함을 보여야 하며 모른다고 인정하는 것을 두려워하지 말아야 한다. 그들은 과학적 방법의 아버지인 프랜시스 베이컨의 조언을 들어야 한다. "사람이 확신으로 시작하면 의심으로 끝날 것이다. 그러나 의심으로 시작하면 확신으로 끝날 것이다."

말만으로는 할 수 없다는 것을 인식하라

궁극적으로 실험 중심 조직의 리더가 된다는 것은 직원들에게 스스로 테스트를 수행할 수 있도록 권한을 부여하는 것을 의미한다. 이는 사람들에게 단순히 그렇게 할 수 있다고 말하는 것만으로는 불가능하다. IBM 과 같은 공동의 노력이 필요하다.

Booking.com 의 사이트 개선 실험 방법

회사의 직원들은 매일 A/B 테스트를 통해 조정 아이디어를 시험해 본다. 다음은 두 가지 예이다.



2015 년 실험은 IBM 에서 핵심 활동이 아니었다. 회사의 IT 부서에서 테스트를 실행하겠다고 제안했지만 비용이 많이 들고, 관련 사업부에 청구되었으며 엄격한 프로세스를 따라야 했다. 검사팀은단 한 명의 전문가로 구성되었는데, 그 전문가는 게이트 키퍼 역할이었고, 그는 실험이 불필요하다고 느꼈기 때문에 제안된 많은 실험을 거부했다. 그 결과 IBM 은 그 해에 단 97 건의 테스트만 실행했다.

그 후, 당시 IBM 의 마케팅 분석 책임자였던 아리 셰인킨이 실험을 맡았고, 최고 마케팅 책임자의 지원을 받아 전 세계 5,500 명이 넘는 마케터가 직접 테스트를 실시할 수 있도록

권한을 부여했다. 세인킨은 마케터가 테스트를 실시하도록 유도하기 위해 여러 가지 조치를 취했다. 그는 사용하기 쉬운 도구를 설치하고, 지원을 제공하는 센터를 만들고, 규율 있는 실험을 실시하기 위한 프레임워크를 도입하고, 모든 사람에게 교육을 제공하고, 모든 사업 그룹에 온라인 테스트를 무료로 제공했다. 그는 또한 마케팅 부서가 30 일 동안 총 30 개의 온라인 실험을 실행해야 하는 초기 "테스팅 블리츠(testing blitz, (역자주)소프트웨어 개발 및 품질 보증 과정에서 사용되는 용어로, 짧은 시간 내에 집중적으로 테스트를 수행하는 것) "를 실시했다. 그 후 그는 가장 혁신적이거나 가장 확장 가능한 실험을 위한 분기별 경연 대회를 열었고 또한 보다 강력한 전술을 사용했다. IBM 은 마케팅 부서 예산의 일부를 실험 계획에 묶었다. 이러한 노력은 효과가 있었고 2018 년까지 연간 테스트 수는 2,822 개로 급증했다.

실험의 변혁적 힘을 실현하려면 지속적인 헌신이 필요하다. 시간이 지남에 따라 실험은 수천 개의 작지 않은 변화를 초래하여 집단적으로 엄청난 이익을 창출한다. 필수적이기는 하지만 적절한 도구를 제공하는 것은 쉬운 부분이며 실험을 삶의 방식으로 만들기에 충분하지 않다. 비스만스는 이를 가장 잘 표현했다. "CEO 에게 조언이 있다면, 대규모 테스트는 기술적인 것이 아니라 온전히 받아들여야 할 문화적 문제입니다. 두 가지 큰 질문을 스스로에게 해야 합니다. 매일 얼마나 틀렸는지 직면할 의지가 있습니까? 그리고 당신을 위해 일하는 사람들에게 얼마나 많은 자율권을 줄 의향이 있습니까? 그리고 답이 당신이 틀렸다는 것이 증명되는 것을 좋아하지 않고 직원들이 제품의 미래를 결정하기를 원하지 않는다면, 그것은 효과가 없을 것입니다. 당신은 결코 실험의 모든 이점을 얻을 수 없을 것입니다."

교훈은 어떤 한 실험이 성공하든 실패하든 그렇게 중요하지 않다는 것이다. 중요한 것은 조직의 불확실성 속에서 어떻게 결정을 내리는 가이다. 그것은 신앙이나 개인적인 의견에만 의존해서는 안 된다는 것이다. 시험할 수 있다면, 시험해야 한다. (끝)

번역자 약력:

- 고려대학교, 미국 알래스카주립대(석사) 졸업
- 현) 이노베이션포럼 수석부회장
- 현) ISO/ TC279 혁신경영전문위원회 위원장
- 현) 이노베이션 프레임워크 대표이사
- 전) 한양대학교 경영전문대학원 겸임교수