****

9B17A053

beworks: experimentation in business

Katie Chen wrote this case under the supervision of Professor Neil Bendle solely to provide material for class discussion. The authors do not intend to illustrate either effective or ineffective handling of a managerial situation. The authors may have disguised certain names and other identifying information to protect confidentiality.

*This publication may not be transmitted, photocopied, digitized, or otherwise reproduced in any form or by any means without the permission of the copyright holder. Reproduction of this material is not covered under authorization by any reproduction rights organization. To order copies or request permission to reproduce materials, contact Ivey Publishing, Ivey Business School, Western University, London, Ontario, Canada, N6G 0N1; (t) 519.661.3208; (e)* [*cases@ivey.ca*](mailto:cases@ivey.ca)*;* [*www.iveycases.com*](http://www.iveycases.com)*.*

Copyright © 2017, Richard Ivey School of Business Foundation Version: 2017-10-26

In 2016, Ada Le, an associate at Toronto, Ontario-based management consulting firm BEworks Inc. (BEworks), sat at her desk on a Monday afternoon. She had just left a meeting with BEworks’s latest client, a luxury cosmetic company. The client had approached BEworks to gain a better understanding of its customers in an effort to develop better-tailored communications. Of special interest was the design of key messages to drive sales growth. BEworks offered a unique approach to strategic problems by using empirical data obtained through experiments, and the client was curious to see this methodology in action. As a result, the client and BEworks had together decided on a variety of promising messages to assess. It was now up to Le to decide on the design of a test that would best gauge the effectiveness of these messages.

The role of testing in business

In the world of business, data-driven decisions were critical to success. However, the data used to support business decisions were not always accurate. Imagine that this year, a clothing company redesigned its logo in an attempt to appeal to more consumers. The company executives may have compared this year’s sales with the previous year’s sales to identify the impact of its new logo. However, by doing so, the business failed to account for extraneous variables that could have also influenced the result. Perhaps demand for clothing simply increased this year, or a key competitor recently went bankrupt. Businesses often made changes to their products and compared the resulting data against historical data to understand the degree of success or failure. While comparing year-over-year data could be useful, it could mislead businesses into making the wrong conclusions.

Testing was pivotal to gathering primary data, which could then be used as evidence to support business decisions. That is, an executive did not simply assume what actions would result in the desired outcome but supported any assertion with evidence. The best evidence came from experiments—scientific tests designed to measure the potential outcome of any test condition. In any situation, the conditions were the versions of product characteristics under consideration for change, such as different flavours or packaging designs.

By developing and conducting experiments, businesses could better understand the true impact of any proposed change. For BEworks, helping its latest client in terms of client engagement meant conducting an experiment. This would allow the client to gain a greater understanding of which message performed best before planning an expensive rollout. Armed with the experimental data, BEworks could deliver recommendations to its client that were backed by empirical evidence.

BEworks

Founded in 2010, BEworks was the world’s first management consulting firm that specialized in applying behavioural economics to real-world challenges. Behavioural economics was an emerging field that sought to understand human behaviour in the context of psychological, social, cognitive, and emotional influences. A typical behavioural economics approach might apply insights from psychology to economic and business problems. A classic example involved finding the best use of social cues to encourage more environmentally-friendly behaviour.

The firm’s co-founding team included Dan Ariely, a well-known professor on the marketing faculty at Duke University. Ariely wrote *Predictably Irrational*,[[1]](#footnote-1) a popular book that introduced behavioural economics to the public. Ariely wasn’t the only scientist on the team; additional research credentials were brought by Nina Mažar, a marketing professor at the University of Toronto’s Rotman School of Management and a behavioural scientist at the World Bank. Mažar was known for her work on understanding the psychological foundations of dishonest behaviour. A third founding partner was Kelly Peters, an experienced strategy and innovation executive who, in 2008, launched one of the world’s first behavioural economics practices at a major bank. She designed the methodology that underpinned BEworks’s practice of the scientific method with behavioural insights and management consulting.

BEworks’s mission was to apply scientific thinking to marketing and operational challenges through an emphasis on empirical testing and data-driven decisions. The firm worked with clients in a variety of sectors, including energy, financial services, consumer-packaged goods, insurance, and government.

As stated on the BEworks website, “Our team of scientists and business experts offers a powerful methodology that analyzes and measurably influences the decisions consumers make. We offer a systematic approach to solving business challenges.”[[2]](#footnote-2)

When working with its clients, BEworks applied a unique methodology, the “BE Method,” which was divided into five steps:

1. Discovery—studying the challenge from a scientific perspective, which included setting the objectives and specifying measurable targets.
2. Behavioural Diagnostics—understanding the client’s challenge through a behavioural lens, using findings from academic and managerial research.
3. Ideation—identifying and prioritizing behavioural economics strategies as potential solutions. In this process, BEworks often leveraged its own “Expert Think Tank.”
4. Build and Experiment—designing and launching experiments to test different solutions, and analyzing the results to provide solutions empirically validated by experiments.
5. Choice Architecture—delivering the “Choice Architecture Playbook,” which included final recommendations based on experimental data and supporting a large-scale market rollout.

**Overview of the cosmetics industry**

The cosmetics industry manufactured a wide variety of products, including makeup, lotions, facial and hair products, and other toiletries. The industry covered a wide range of price and quality points from inexpensive pharmacy brands to luxury brands such as Estée Lauder and L’Oreal Paris. Cosmetic products were often discretionary purchases; for example, when money was tight, spending on cosmetics could be more easily decreased, compared with spending on food. The performance of the industry was thus closely related to the Consumer Confidence Index, which tracked consumer sentiments on the state of the economy. As a result of uncertain consumer sentiments, the industry faced a projected negative annual growth of −0.5 per cent though to 2021.[[3]](#footnote-3)

The advent of technology and e-commerce had significantly reduced the barriers to entry, allowing smaller companies to access and take advantage of niche markets. Given the increasing ease of market entry, the number of players in the market was expected to increase at an annualized rate of 4.1 per cent over the next four years to 2021.[[4]](#footnote-4)

The cosmetics industry could be categorized into low-end and high-end manufacturers. The low-end manufacturers competed more on price, while the high-end, or luxury, manufacturers, competed on quality. Given the competitive pressures, manufacturers also needed to consistently invest in research and development (R&D) to stay relevant to the consumer by offering new, innovative products. Finally, the industry was also affected by several significant social trends, such as the popularity of organic and all-natural beauty products.

The client’s challenge

The luxury cosmetics client had approached BEworks to gain insight on how to improve product messaging on one of its skincare lines. The client wanted to understand what message would best generate an understanding of the product’s value proposition, and as a result, drive sales of its products.

The BEworks team and the client had decided on a total of 17 different product taglines to test (see Exhibit 1). Each tagline had been generated through discussions between the client and BEworks. Ideas from behavioural economics literature and managerial experience suggested that these specific taglines might be powerful in engaging consumer interest.

The messages were categorized into six themes, with the first five categories being Quality, Custom Fit, Science, Time Horizon, and Whole Unit Framing messages. The first five themes sought to convey important consumer values identified through prior consumer research:

* Quality messages aimed to convey that the client’s products were high-performing and provided great results.
* Custom fit messages aimed to convey that the client’s product line could be personalized.
* Science messages emphasized all the R&D efforts that have gone into the products and linked the client’s dermatology expertise to the products.
* Time horizon messages emphasized the product’s ability to achieve results within a short time frame.
* Whole unit framing messages aimed to frame the entire product line as a bundle to encourage the purchase of the entire product line.

The final category of themes, control messages, consisted of taglines that were either currently deployed on the client’s displays or had been used frequently in the past. These control messages would be tested alongside the new messages, with their results serving as a benchmark against the results of the newly proposed messages. The data generated from the test would allow BEworks to answer a question critical to any company planning to invest in marketing: Are the new taglines any better than those currently in use?

Designing the test

Before BEworks could provide data-driven recommendations, it needed to first generate reliable data, and the best way to do so was by conducting an experiment. One of the cornerstones of experiments was randomization. The idea behind randomization was simple: by maximizing the probability that the groups receiving each condition were similar, any observed differences between groups could therefore be attributed to the impact of the condition itself. If a researcher tried one tagline out on younger women and another one on older women and found that one was more effective, the researcher would never know whether it was the tagline that drove the difference or the age of the women. By using random assignment in an experiment, and by having a large enough number of subjects in each condition, the laws of probability suggested that the groups would be relatively comparable in terms of demographic and other relevant variables. For example, the average age of women in each group should be roughly the same, effectively removing age as an explanation for any differences.

Le knew that randomization was incredibly important to the test that BEworks would conduct. To carry out the test, Le was deliberating among three alternatives: a laboratory experiment, a field test without randomization, or an online survey. She wanted to carefully consider the pros and cons of each approach and identify which option would be the best fit to test proposed product taglines for her client.

A Laboratory Experiment

Laboratory experiments were conducted in a laboratory—a heavily controlled environment. Subjects were randomly assigned to one of several conditions, and the results of subjects across different conditions were then compared against each other and any control conditions. For Le’s experiment, the control conditions would be the status quo treatment (i.e., the current treatment). If Le were to conduct a laboratory experiment, each tagline would have one condition, for a total of 17 conditions: 13 experiment conditions and four control conditions.

Laboratory experiments minimized the impact of external variables, which were extraneous factors that could also influence the measured result. For example, if a laboratory wanted to measure the impact of a particular stimulus on heart rate, it could hold the temperature constant for all subjects, therefore reducing the potential impact of different temperatures on the heart rate measurement. However, this artificial control could also be a drawback when testing any conditions where external variables would always exert influence in the real world. For businesses, many consumer-purchasing decisions were driven by social factors—such as seeing other shoppers purchase the product. A laboratory experiment would typically remove the effects of variables such as social influence, but for businesses, such lab results may not be indicative of real-world outcomes. Le wondered whether a laboratory experiment would be a good fit for her client’s goals.

Field Testing without Randomization

Field testing was performed in the environment where the behaviour or interaction of interest took place naturally. For example, for a retail study, the test might take place in a store. As such, the results generated would be strong predictors of actual outcomes. However, randomization was much harder to achieve in the field—a real-world environment. For example, it was typically impossible to offer different prices, packaging, or taglines to randomly selected shoppers in a brick-and-mortar store.

One option was to run a test that violated the randomization requirement of a true experiment—subjects were not randomly assigned to conditions. Using natural settings for tests that did not achieve perfect randomization could still be beneficial to businesses but would not provide the same control as provided by laboratory experiments. For instance, external variables such as weather would vary between customers in different locations, which might affect sales if variants were being tested in different geographical regions. In those circumstances, statistical corrections could, to some degree, compensate for the limitations, which meant an advanced understanding of statistics was important to conduct many of the more advanced field tests where randomization was imperfect.

If Le were to use a field test but not seek to randomize, the different messages would be physically tested, in the field, among shoppers across different geographical regions. Perhaps one tagline could be used in one group of stores and another tagline could be used in another similar group of stores, with Le comparing sales across the different groups. Although this approach would most closely indicate results in the real world, she would need to consider the feasibility of this approach and the areas where errors could potentially occur.

An Online Survey

The third option, an online survey, would be a relatively practical method to collect data for BEworks’s client. Electronic surveys were incredibly cost-efficient, time-efficient, and easy to deploy. Amazon Mechanical Turk was the best-known method of conducting such surveys; however, more specialized providers, such as Qualtrics, provided more-customized solutions. The considerations regarding each approach varied. For example, Mechanical Turk workers often conducted a large number of surveys in a short span of time for little compensation.

The technology available made it extremely easy to randomly assign respondents to different conditions, which could allow for more sophisticated approaches than available with many online surveys. This online panel experiment approach was not the same as the traditional approach of customer satisfaction surveys that simply ask respondents to give their assessments of the products based on their experience, with respondents using the same survey. It also differed from asking respondents to think about and give their opinions on a range of messages. Rather than presenting all the messages to all respondents, only one message would be presented to each individual in the online panel experiment. This approach allowed for differences in response to be seen for each condition. Given the random assignment, the differences in response could be confidently attributed to the message seen.

However, Le had some concerns about conducting an experiment online. Because consumers would simply be assessing different product taglines only from computer screens, their answers would be predictive but not necessarily fully representative of their actual behaviour when confronted with the product and associated messaging in-store. The data collected could be only a first stage to estimate the impact. If Le chose to use an online survey, could an online experiment be part of a wider testing strategy? What testing would follow next?

If Le were to choose the online survey, she would need to consider the target demographics. What geographic location, gender, age, and household income should she target for the survey? Were there additional demographic criteria she should also consider?

Conclusion

BEworks’s engagement with this client was on a tight timeline, so Le did not have much time to decide. As she started typing on her laptop, she knew that her team was depending on her to prepare a slide deck presentation of the benefits and drawbacks of each experiment design and a recommendation on which experiment design would be best for the client. This project could have a huge impact on the client’s position in the competitive luxury cosmetics market.

Exhibit 1: Messages for Testing

|  |  |
| --- | --- |
| **Theme** | **Message** |
| Quality | High performance skin care to give you results you can feel. |
| We’ve helped millions of women find their way back to skin that’s so clean, it glows. |
|  |
| Custom Fit | 64 combinations. Only one that fits your skin. |
| Our formula to fit your skin. |
| Customized care for beauty that’s all yours. |
| A skin care system that’s all yours. |
|  |
| Science | No bells. No whistles. Just science-driven results. |
| Beauty is an art. Skin care is a science. |
|  |
| Time Horizon | There’s great skin in your future. The future starts now. |
| Happier skin tomorrow. |
|  |
| Whole Unit Framing | The three must-haves for irresistibly smooth skin that is uniquely yours. |
| Three simple steps, two times a day, one happy you. |
| Lather, tingle, quench: Feel the experience. |
|  |
| Control Messages | What is the price for great skin? |
| Great skin can be created. |
| Get great skin today. |
| Great skin. Or your money back. |

Source: Company documents.

1. Dan Ariely, *Predictably Irrational, Revised and Expanded Edition: The Hidden Forces That Shape Our Decisions* (New York, NY: Harper Collins, 2010). [↑](#footnote-ref-1)
2. “Our Methodology: What Is Behavioral Economics?,” BEWorks, 2017, accessed March 8, 2017, http://beworks.com/Our-Method. [↑](#footnote-ref-2)
3. “Industry Reports (US): Cosmetic & Beauty Products Manufacturing in the US—32562,” IBISWorld: Where Knowledge is Power, May 2017, accessed August 1, 2017, www.ibisworld.com/search/default.aspx?st=cosmetic%20and%20beauty%20products. [↑](#footnote-ref-3)
4. Ibid. [↑](#footnote-ref-4)