

Chapter 7

The Ethical Implications of A/B and Multivariate E-Commerce Optimization Testing

J.J. Sylvia IV

The University of Southern Mississippi, USA

ABSTRACT

A/B and multivariate website optimization may not seem ethically problematic at first blush; however, in this chapter I will consider some of the less obvious elements that have been tested, such as header color, button design, and the style of tabs used for linking to product details. A/B and multivariate testing has shown that these seemingly insignificant changes can increase average order value and decrease abandoned shopping carts, among other results. I will consider these tests through the lens of the major ethical systems of utilitarianism, Kant's respect for person's principle, and virtue ethics, using specific case studies and examples of testing results. I conclude that this type of practice is likely ethically problematic in many uses, as understood through all three ethical systems. Along the way I will be careful to demonstrate how the manipulation resulting from A/B and multivariate testing is different and more problematic than that of advertising in general.

INTRODUCTION

Certainly, many ethical issues related to the creation and use of websites have already been broached. The 2000 Children's Online Privacy Protection Act addresses the privacy of minors (COPPA, 2009). In addition to privacy concerns, copyright issues have also played a large role in the development of the Internet (Montecino, 1996). A more recent

issue is accessibility standards for web surfers with disabilities (*United States Access Board*, 2009). Some of these issues have been addressed by acts of Congress, while others have gotten their precedent from law suits, but overall, many of these issues are still in flux and still under debate. I will broach an issue that has not been previously addressed from an ethical or legal view.

Manipulative advertisements have long been the focus of major discussions within business ethics, and this discussion can certainly be ported

DOI: 10.4018/978-1-61520-615-5.ch007

over to advertising via the Internet. However, the Internet also opens up a new avenue for manipulation based on the layout, design, and copy of a particular website. It may be possible that many of the elements on a company's website might actually still fall under the topic of advertising; however, many of the changes that are made on websites due to A/B and multivariate testing clearly should not be understood as elements of advertising. For example, one would not be likely to categorize the color and/or style selection of category tabs as a form of advertising. Therefore, although the discussion may sometimes mirror that of manipulative advertising, a new discussion about implementing the results from A/B and multivariate tests needs to be highlighted.

By and large, companies which advertise their ability to implement multivariate tests on behalf of another company promote the activity in a way such that it appears altruistic or morally exemplar, rather than ethically questionable or problematic. Multivariate testing for websites has been occurring since at least 1995, but these tests have recently gained widespread exposure to a much larger audience of web developers. A/B and multivariate testing is rising in popularity of use and increasing in simplicity of implementation. Despite this increased use, the ethical implications of using such tests have not been raised or explored in either professional literature or on the web in general. For example, is it acceptable that websites conduct this kind of testing? Questions such as this have not yet been broached. In this chapter I strive to set the agenda for the broader topic of the ethical implications of multivariate tests, hoping to open up the issue for further commentary rather than trying to cover every aspect in depth.

BACKGROUND

A/B and multivariate tests on the Internet allow a website to test two or more versions of the same

page and measure desired outcomes. A/B split testing allows one to:

randomly divide your visitors into two groups and show each group a different version of a page to determine which version leads to higher conversion, average order value, application completion, or other target. These visitors are then tracked and a report is generated that describes the impact of the A or B page version on this outcome. (Roche, 2004)

Multivariate testing, on the other hand, is:

a process by which more than one component of a website may be tested in a live environment. It can be thought of in simple terms as numerous split tests or A/B tests performed on one page at the same time. (Search Engine Marketing, 2009)

One example of this practice would be a site testing the placement of a search box on the page in order to see if it gets used more frequently depending on whether it is on the top right or left of a page. This type of testing came about long before the Internet and was used for many different purposes; however, the Internet allows for easy use and quick testing of many different design elements at the same time. Results can also be quickly linked to sales related figures such as percentage of shopping carts abandoned or average order size.

Discussion of manipulative advertising has a long history within the realm of business ethics, but this issue falls outside of the typical realm of those debates. The tests I discuss have nothing to do with the products, or brands, or marketing messages themselves. Instead, they focus on elements such as the placement of search boxes, header color, button design, and the style of tabs used. A standard example of manipulative advertising would be a commercial for a fast food company displaying large images of juicy hamburgers that create a feeling of hunger or desire

over to advertising via the Internet. However, the Internet also opens up a new avenue for manipulation based on the layout, design, and copy of a particular website. It may be possible that many of the elements on a company's website might actually still fall under the topic of advertising; however, many of the changes that are made on websites due to A/B and multivariate testing clearly should not be understood as elements of advertising. For example, one would not be likely to categorize the color and/or style selection of category tabs as a form of advertising. Therefore, although the discussion may sometimes mirror that of manipulative advertising, a new discussion about implementing the results from A/B and multivariate tests needs to be highlighted.

By and large, companies which advertise their ability to implement multivariate tests on behalf of another company promote the activity in a way such that it appears altruistic or morally unobjectionable, rather than ethically questionable or problematic. Multivariate testing for websites has been occurring since at least 1995, but these tests have recently gained widespread exposure to a much larger audience of web developers. A/B and multivariate testing is rising in popularity of use and increasing in simplicity of implementation. Despite this increased use, the ethical implications of using such tests have not been raised or explored in either professional literature or on the web in general. For example, is it acceptable that websites conduct this kind of testing? Questions such as this have not yet been broached. In this chapter I strive to set the agenda for the broader topic of the ethical implications of multivariate tests, hoping to open up the issue for further commentary rather than trying to cover every aspect in depth.

BACKGROUND

A/B and multivariate tests on the Internet allow a website to test two or more versions of the same

page and measure desired outcomes. A/B split testing allows one to:

randomly divide your visitors into two groups and show each group a different version of a page to determine which version leads to higher conversion, average order value, application completion, or other target. These visitors are then tracked and a report is generated that describes the impact of the A or B page version on this outcome. (Roche, 2004)

Multivariate testing, on the other hand, is:

a process by which more than one component of a website may be tested in a live environment. It can be thought of in simple terms as numerous split tests or A/B tests performed on one page at the same time. (Search Engine Marketing, 2009)

One example of this practice would be a site testing the placement of a search box on the page in order to see if it gets used more frequently depending on whether it is on the top right or left of a page. This type of testing came about long before the Internet and was used for many different purposes; however, the Internet allows for easy use and quick testing of many different design elements at the same time. Results can also be quickly linked to sales related figures such as percentage of shopping carts abandoned or average order size.

Discussion of manipulative advertising has a long history within the realm of business ethics, but this issue falls outside of the typical realm of those debates. The tests I discuss have nothing to do with the products, or brands, or marketing messages themselves. Instead, they focus on elements such as the placement of search boxes, header color, button design, and the style of tabs used. A standard example of manipulative advertising would be a commercial for a fast food company displaying large images of juicy hamburgers that create a feeling of hunger or desire

for a hamburger in the viewer, which he or she did not previously have. It is not immediately obvious how web page elements such as search boxes and tabs are similarly manipulative. Multivariate test results can make the connection more obvious, and the results could possibly be linked to specific users or demographics. For example, a particular demographic may be more likely to make a purchase when a headline is displayed in a specific color, and could then be targeted by a dynamically generated website which identifies the demographic of the user and displays the color that has been shown to increase the odds of that person making a purchase.

Privacy is another major ethical concern for e-commerce sites. Technology allows a site to record and store a great deal of information about particular site visitors; however, the goal of multivariate testing is not necessarily to tie the results to any particular user. Instead, overall changes in the actions of the majority of users are of concern, so individual privacy is not the main issue. Yet, multivariate testing on the Internet is still quite young, so although individual privacy is not yet of concern with these tests, this practice may still develop in the future and be cause for even further concern. In general, though, multivariate testing falls outside the scope of some of the largest e-commerce ethical concerns, and very little has been written which directly relates to the practice.

THE ETHICAL ISSUE

The ethical issues surrounding the use of A/B and multivariate website optimization testing are important because these tests are now easily available to all e-commerce sites through Google's free WebAnalytics software. One can visit http://www.google.com/analytics/sign_up.html to create an account and have code automatically generated which can be inserted into a webpage. Previously, a great degree of planning and technical expertise was needed to implement such a test, but it has

quickly become both free and easy to implement and thus widely available to any e-commerce site, regardless of its size or budget. This type of testing could have benefited from an ethical analysis at any point, but as the popularity of this type of web development increases, the importance of such an analysis increases. A 2007 Internet Retailer magazine survey of 243 chain retailers found that 23.3% of websites were currently using A/B or multivariate testing (Brohan, 2007). However, this use is certainly on the rise. A 2009 survey of 650 chief marketing officers found that 28% of the marketers planned to deploy multivariate testing this year (*Marketers plan to spend*, 2009). Many companies offer multivariate testing services, but Google alone highlights case studies of companies such as RE/MAX, The Huffington Post, and The American Cancer Society (*Google Analytics*, 2009).

I will argue that this practice leads to manipulative site design aimed at increasing the amount of money spent by consumers on an average order; however, a more charitable interpretation of this testing is certainly available. One could understand the tests as aiming to help consumers best achieve what it is they would like to achieve. For instance, the placement of the search box might be important because a consumer might not be as likely to locate it in one spot rather than another. If the consumer ultimately would like to make a purchase, but simply is not able to because he or she overlooks the search box, it would certainly make sense to test locations of the search box for visibility.

Companies which offer multivariate testing often promote the service as beneficial to not only the e-commerce company, but also the consumers using the site:

Here at CSN Stores, customers are our highest priority and multivariate testing lets us tune our sites so that they are highly intuitive for the largest number of customers," noted Chuck Casto, Vice President of Corporate Communications for CSN

Stores. "By continuously testing and targeting, we are able to enhance the CSN online shopping experience for consumers. (Online Businesses Doing, 2009)

An interpretation such as this suggests that site optimization simply makes it easier for a consumer to complete some task he or she was already trying to complete. Certainly there are situations in which this really is the case, but clearly not all tests are aimed at such a lofty goal. Instead, the emphasis is often on *creating* a conversion or sale.

On a surface reading, this does not seem any more problematic than a brick and mortar store putting candy bars next to the checkout or playing carefully selected music throughout the store. However, multivariate testing via the Internet allows for much faster and more accurate results than other media. One is able to draw direct statistical conclusions from the changes made on a webpage to changes in sales figures. Why does the design of a button or the color of a heading affect the likelihood of a person completing a purchase? As of yet, not many tests have been very concerned with this 'why' question because the focus is on the results. Without a solid answer to the 'why' question, though, the practice of multivariate testing can be understood as an extremely manipulative practice. By implementing the results of these tests, web developers can be understood as using minute, manipulative tactics in order to increase the amount of money spent by consumers. Michael Phillips defines manipulative advertising as "advertising that tries to favorably alter consumers' perceptions of the advertised product by appeals to factors other than the product's physical attributes and functional performance," (1994). I believe this definition can be adjusted for manipulative web site design in the following way: manipulative design is design which tries to increase the likelihood of a customer purchase or the amount of a purchase by altering design elements unrelated to information about a product's physical attributes or functional performance,

using design changes that have been statistically shown to lead to such increases.

At least part of the answer to the 'why' question will likely lie with psychological responses. For example, particular colors are often hypothesized to create a particular mood or emotion in the person experiencing that color. The employees of Google have done research on the effect of the shade of blue used in links:

Google Mail uses a very slightly different blue for links than the main search page. Its engineers wondered: would that change the ratio of click-throughs? Is there an "ideal" blue that encourages clicks? To find out, incoming users were randomly assigned between 40 different shades of links – from blue-with-green-ish to blue-with-blue-ish. It turned out blue-ness encouraged clicks more than green-ness. Who would have guessed? And who would have cared? Google, of course, which wants to get people clicking around the net. (Arthur, 2009)

For a detailed critical review of empirical research on the affects of color, reference Whitfield, T., & Wiltshire, T. (1990).

In addition to color, there is perhaps a similar reaction to other elements in web page layout and design. Consider example results from ShopNBC.com:

The five tested elements were the location and size of the Add to Cart button; headlines identifying cross sells; styles of the tab linking users to product detail; color, size and style of clearance and limited time pricing offers; and highlighting of payment options for qualified buyers.

ShopNBC found that variations in a single element with the others remaining unchanged produced a relatively small effect in improving results. The right combination of options drawn from all five elements, however, produced the 16% lift in aver-

age order value, besides increasing the number of shoppers who initiated the checkout process by 4%. (*Multivariate Testing Produces, 2008*)

Results such as this are not atypical. The changes produced only a small result in the number of consumers actually initializing a checkout. Instead, what increased was the average order value. The main problem was not that many customers were unable to figure out how to add items to their cart or checkout. The changes simply influenced them to buy *more*. Again, no one seems to care about *why* the style of a tab or the color of a clearance offer increases average orders—the pragmatic results are enough. Manipulation of the consumer, that is, making the customer more likely to buy a product or products through layout and design changes unrelated to the details of the merchandise being sold, is often viewed as the desirable end goal rather than a potentially ethically problematic practice. Companies are actively being encouraged to engage in these practices in order to increase profits. This encouragement has thus far lacked any ethical reflection on the potential problems associated with such practices.

Contrast this with the oft-considered ethical issues of manipulative advertising. One issue with advertising is that it is said to *create* desires in consumers that did not exist before. The process is relatively straight forward and easy to understand. I see an advertisement with a giant picture of a hamburger and this makes me desire a hamburger, even if I had not been feeling particularly hungry before. Whether or not this practice is unethical can be debated in a somewhat straightforward manner. On the other hand, it is not nearly as clear what is going on as the result of multivariate tests. How could the style of tabs and the size of a checkout button possibly influence one to purchase more?

Perhaps a better analog to an ethical issue is the use of color in physical stores and logo designs. Many restaurants prominently feature the color red in their stores and logos because the color has been shown to induce hunger in those seeing

it (Stoll, 2004). For a similar reason, check-out lines in grocery stores are lined with candy in the hopes of inducing an impulse purchase. What is different about online multivariate testing is the ability to test several layouts at the exact same time while continuously monitoring information. If a particular change produces the best results only on weekends, for example, while another produces the best results on weekdays, little effort is required to display the optimal page layouts on different days, at least compared to the effort of reorganizing a grocery store to have different displays on weekends than weekdays. Thus, A/B and multivariate testing issues lie somewhere between the ethical debates relating to advertising and marketing.

The problem is that there has to be some type of design on a site and an easy to navigate, functional interface is certainly desirable. The difficult question that remains is: what is the difference between acceptable good design practices and manipulation? The answer may be intent. A site which is created with the goal of being user-friendly and aesthetically pleasing seems laudable; however, once design changes are made solely on the basis of how much extra money these changes can pull out of consumers, a line is perhaps crossed. Instead of the intention being to create a usable site, the intention shifts to doing whatever is necessary to make more money, and this includes mental manipulation of consumers. If seeing a headline in red rather than blue can be shown to statistically generate more revenue, then by all means the site will post their headlines in red.

Another aspect of the difficulty in dealing with the issue is the phenomenological experience on both sides. As someone shopping on the Internet, I do not directly experience subtle elements such as colors or styles pulling me toward some purchasing decision. On the other end, I have also worked as a web developer; although I never used multivariate testing specifically, one of the goals was to design the site in such a way that it would maximize profits. For example, adding a feature

which allows for magnification of an image in order to better see the details seems like a good idea that will help consumers decide whether or not they like a particular product. It may even be the case that in seeing the detail, they decide this is a product that they do not want. The phenomenological experience of making a change such as this and making a change in the tab style or button color is very similar. Although the actions in these two situations are similar, the goals are completely different. In the former example, the goal is to help the consumer gain more information about a product and enable them to better determine whether or not he or she really desires to make a purchase. In the latter case, the goal is to increase the chance that the consumer will want to make a purchase or buy more items, regardless of whether or not he or she really desires to make the purchase. For a web developer who must necessarily be worried about increasing profits, taking the time to understand and reflect on the differences in these somewhat similar situations may be quite difficult.

Further, with corporate focus heavily emphasizing bottom lines and profit increases, web developers may feel pressure to make any change that is going to increase sales, regardless of the reason or manipulation involved. On one hand, these design tweaks aimed at manipulation might be easily ignored in a free market economy where "*caveat emptor*" is still a predominate sentiment. Certainly it seems unlikely such actions would ever be deemed illegal, yet it seems pertinent to consider whether such actions might be unethical. In trying to analyze the ethical implications of A/B and multivariate testing, I will be making several basic assumptions about which I would like to be clear up front. I will assume that based on the statistical results of multivariate tests, it is the actual changes to site design that are driving the increase in sales and not some outside force or circumstance that was not able to be measured. I believe enough case studies have been completed to show that these sometimes minute

changes can make a difference in sales. I will consider whether there may be ethical problems with such an approach through three ethical systems: utilitarianism, Immanuel Kant's deontology, and virtue ethics. I have selected these systems because they are the most common ethical views typically discussed in relation to business ethics in general. Within each, I will assume the validity of the ethical system being discussed, in order that criticisms of a particular ethical system should not be lodged against the ethical analysis made through that particular system.

Utilitarianism

Utilitarianism is a system based on a Greatest Happiness Principle which states that "actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness," (Mill, 2004, p.6). Further, John Stuart Mill defines happiness to be pleasure and the absence of pain. This system of ethics was founded by Jeremy Bentham but greatly expanded by John Stuart Mill. Although many variations of utilitarianism have since arisen, I will be considering this issue mostly from the perspective of hedonistic act utilitarianism, meaning the consequences of each individual act should be considered in making a moral judgment.

A utilitarian must ask if the overall good is served by manipulating consumers into buying more. Certainly, on the side of the company this does seem to be toward the good. Similar utilitarian arguments have been made supporting the use of manipulative advertising, such as the statement by William H. Genge: "Where does the money go? The answer is: It provides jobs and livelihoods for hundreds of thousands of people – not only in the advertising and communications sector but for all the people employed by fast-food companies and, indeed, all marketing organizations" (1985, pp. 58-59). Genge is arguing that advertising money is not wasted because the advertising causes consumers to make more purchases, and these

purchases keep the economy moving and growing, allowing for more people to be employed – thus more people make more money and presumably the overall good is increased.

A similar argument could be applied to the manipulative changes made to web pages due to multivariate testing. Although the consumer may be manipulated to purchase more than he or she would have on his own, this is to the positive because the extra money allows the company to continue growing and expanding, possibly even hiring more employees as they do so and thus increasing wealth for many people. Sales allow for the economy to continue moving. What of the consumer, though?

The negative impact on the consumer does not *prima facie* seem to be very large. He or she was presumably already seeking to make some kind of purchase, and although the amount of the purchase has increased, that increase by itself is not typically significant. Average orders are increasing in the range of ten to twenty percent not two or three hundred percent. Although there may be a higher likelihood of buyer's remorse, the negative impact hardly seems to be problematically large. On this reading, many people see an increase in utility as the economy grows, while a consumer sees a small decrease in utility. Perhaps multivariate testing is not problematic from a utilitarian perspective.

Along those lines, arguments have been advanced that advertising and purchasing products actually increase overall happiness. In summarizing such an argument by Theodore Levitt, Michael Phillips says that "we do not merely buy a physical product, but also a set of positive feelings connected with it by advertising. If his [Levitt's] argument is sound, those feelings give us extra utility above and beyond the utility we get from the product's performance of its functions" (1994, p. 41). Although multivariate testing does not create exactly the same type of positive feelings as advertising, it must in some sense create a desire to purchase more, therefore

the consumer might be understood as being able to fulfill a further desire.

The impact of these arguments arises because of the capitalistic assumption that the economy should always continue to grow and that making purchases really does make a person happier. Yet, this basic assumption is rarely questioned within the advertising community. It does seem clear that purchases can and do increase the happiness of a consumer. We have all witnessed and likely experienced the euphoria of making a large purchase and the joy that comes with it – whether it be a new car or a new computer. However, that euphoria very often wears off quite fast. Capitalism provides the answer to this problem as well though, because once the euphoria wears away and the computer starts running more slowly than we would like, we can just buy yet another new computer which brings back that feeling of euphoria in a never-ending cycle. Presumably that cycle of buying and the related euphoria is the good life, and offers increased utility all around. Consumers buy more and more, growing the economy and jobs, and thus everyone can afford to continue buying more and more, bouncing from one euphoric purchase to another.

This is a good story, but I am not entirely convinced that it is the whole story. Yes, that euphoria exists; yes we can keep achieving it by purchasing more and more, but when this becomes the singular focus – when businesses are given the okay to manipulate others to make purchases, many can feel that something is missing. Consider John Stuart Mill's distinction between the quantity and quality of pleasure. In opposition to Jeremy Bentham's strict utility calculations, Mill believed that some pleasures are better than others. If one accepts Mill's distinction, it then becomes an important question to ask what the quality of pleasure derived from making purchases is and whether an alternative option exists that would provide a higher quality of pleasure. This is an open question that has yet to be settled, but I believe a very strong argument could be made that

the pleasure from the euphoria of purchases is of a much lower quality than other activities such as having fun with friends and family. Although the empirical data is not completely cut and dry, research does seem to show that once basic needs are met, increased wealth does not seem to increase happiness (Meyers, 2004). Mill himself believed that one needed to appeal to a "competent judge" in order to determine an issue such as this. In other words, only a person capable of fully experiencing both types of pleasures could truly determine which is the higher pleasure (Mill, 2004, p. 15).

The utilitarian case against multivariate testing is further complicated by the use of such tests by philanthropic organizations. The American Cancer Society used Google Analytics to improve many facets of their web campaign, from search engine advertising copy to homepage layout to email marketing. Alexander Negash, the user experience manager for the American Cancer Society says, "Google Analytics is literally helping in the fight against cancer" (*Google Analytics – Case Studies*, 2009). Even if one understood the American Cancer Society as using multivariate testing with the strictest of intentions to increase the amount of money spent by their visitors – which is clearly not the case here – a utilitarian may still have trouble criticizing such an action because of the overall good being done. Further, the visitor is not being manipulated into being a materialistic consumer – if anything, one would have to understand them as being manipulated into being an altruistic giver.

Should a utilitarian support implementing the manipulative changes suggested by multivariate testing? This remains a bit of an open question, because the results of the actions are somewhat difficult to fully deduce. The typical story is that purchases do increase utility overall, but that story has recently been called into question more and more. It is certainly a possibility that the huge emphasis on purchasing and consumerism is actually not good for a person, but perhaps the better argument is that there are alternative options that are of a higher quality of pleasure. For

example, one might gain pleasure, and thus utility, from buying a new pair of shoes every week; however, a higher quality of pleasure might be found instead in reading a book which has been checked out from one's local library. Although pleasure is to be found in both activities, many would argue that becoming engrossed in a book is of a much higher quality pleasure than buying new shoes. If this can be safely asserted, then a utilitarian should oppose manipulative site design aimed solely at increasing conversion rates on a product like shoes.

To take this analysis one step further, it should be noted that Mill acknowledges the difficulty in applying act utilitarianism to every decision which needs to be made. For that reason he supports a type of rule utilitarianism which says that one should do those actions which, when followed, typically tend to produce the greatest good for the greatest number, even if they might not in every particular case. Looking at multivariate testing from this perspective, one might argue that the practice of making decisions based on emotional manipulation, while beneficial in certain cases, as a rule makes for a less rational and self-disciplined society. Overall, rule utilitarianism could offer a more substantial criticism of such a manipulative practice, similar to virtue ethics which is analyzed later.

Deontology

In his deontological ethical system, Immanuel Kant proposed that one should always treat others never merely as a means, but always also as an end (Kant, 1993, p. 30). In other words, in order to treat a person ethically one must respect his or her autonomy. Drawing from literature on manipulative advertising, one can see that the case has been made that advertising undermines the autonomy of consumers. This same type of argument might be lodged against the implementation of results from multivariate testing. If designing or laying out a page in a certain manner consistently causes

consumers to spend greater amounts of money, perhaps this is an assault on autonomy.

Even within ethical debates on advertising, this point may be disputed, and I believe it may be more difficult to make the same case for multivariate testing. For starters, Andrew Sneddon says one can consider the phenomenological experience of being a consumer: "it typically does not feel like one is being jerked around and parted from one's money like a puppet" (2001, p. 15). I am sympathetic with that reading of manipulative advertising, and even more so when it comes to multivariate testing. It is almost difficult to fathom that when one is shopping on an e-commerce site, something other than one's autonomy is driving the purchase decision.

Advertising's manipulation of autonomy at least has a somewhat simple story behind it: advertisements create or enhance desires and offer products as easy methods of gratification. Almost anyone can consider this view and find it at least within the realm of feasibility that this may actually happen. We all have desires, and certainly different experiences can cause desires to arise. Looking at a picture of a large hamburger can certainly make a person hungry and create a desire for a hamburger that did not exist before.

The effects of multivariate testing do not offer a story that is understood quite so simply. At best one might argue that a psychology of aesthetics is being employed in which the layout, color, and size of various elements is able to manipulate one in a way that they want to or are more comfortable with spending more money. However, this story is not as easily intuited. I can easily think of the time I have seen a hamburger and felt hungry, but there is not a similar experience I can point to where I experienced a red headline and suddenly felt an urge to spend more money.

If the phenomenological experience does not reflect it and the common sense story is hard to interpret, is there really any manipulation going at all? Roger Crisp offers a discussion of manipulation and offers an understanding which might be

able to make sense of what is happening: "A more convincing account of behaviour control would be to claim that it occurs when a person causes another person to act for reasons which the other person could not accept as good or justifiable reasons for the action" (1987 p. 416). Through this interpretation, we might consider something to be a case of manipulation if it causes action due to reasons the person being manipulated would not find justifiable. This offers a clear way of understanding the manipulation going on in the case of multivariate testing. If I were to know that I was going to spend extra money simply because of the aesthetic design of a particular page that had been fine-tuned for just such a purpose, I would not find that to be a justifiable reason for my increased purchases. Therefore, according to Crisp, I could conclude that I had been manipulated into spending more money. This manipulation undermines my autonomy and therefore the website designers could be understood as acting unethically under Kant's respect for person's principle.

Yet, one further problem to this type of argument remains: every web site must have some type of design to it. Even something as simple as using black text on a white background is done by deliberate choice when creating a web page. How can one differentiate between good design principles and unethical manipulation of consumers? An answer to this type of question is not entirely clear.

A possible answer to this problem would be to suggest that any use of A/B or multivariate testing is problematic in that it aims to manipulate consumers. Unfortunately, reality is not that cut and dry. As I have mentioned, there are many cases where legitimate ease of use questions might be resolved by multivariate testing. Perhaps an information site is experiencing a large number of visitors leaving after spending a few seconds on the main page. This may be because the search box is not well placed or is too small to be noticed. A multivariate test may help bring this out, and allow a site visitor to search for the information

they wanted to find. It would be much more difficult to make the case that any type of unethical manipulation is going on in this multivariate test, so the answer must be more complex than claiming that all multivariate tests are unethical.

The intentions of the web developer may be the strongest criterion from which to make an evaluation of the ethical position. What exactly is the goal of the multivariate test? What is being measured? If the main criterion is the increase of income, this may be a warning sign that the intentions of the web developer are unethical.

Aside from multivariate tests, other methods exist for measuring outcomes which may actually be more useful if the sole criterion is not increased profit. Focus groups can use a website or a particular page. In addition to the verbal feedback that they are able to give, their actions on the page and even their eye movement can be tracked. This data can best answer questions about what the person was trying to achieve, whether or not they had trouble, and also what could be done to improve the webpage. In a focus group, the testing is more clearly focused on helping a consumer achieve something they were already attempting to do, rather than searching for any combination of factors that will increase sales. The biggest problem with focus groups as opposed to multivariate testing is the much greater expense and effort involved in setting up a focus group as opposed to a multivariate test.

Drawing the line which demarcates a particular multivariate test as unethical under a deontological system using the respect for person's principle is a difficult task, but I argue that at least in some cases, the intentions of the web developer clearly makes the implementation of the results of such tests an unethical action. Using these results is a form of manipulating consumers.

Virtue Ethics

Whereas deontology is a rule-based system and utilitarianism focuses on the consequences of ac-

tions, virtue ethics considers a person's character to be of chief importance. An Aristotelian interpretation of this would suggest that the best characters are those that experience emotions that are the means of two extremes. For example, a virtuous person acts courageously, rather than rashly or cowardly. Virtue ethics is often thought to offer one of the best arguments against manipulative advertising (Phillips, 1994). In the video *Advertising and the End of the World*, Sut Jhally notes that one of the main purposes of advertisements is to tell us that we can become happy through the consumption of objects (1998). Neil Postman and Steve Powers reiterate this message:

Boredom, anxiety, rejection, fear, envy, sloth – in TV commercials there are remedies for each of these, and more. The remedies are called Scope, Comet, Toyota, Bufferin, Alka-Seltzer, and Budweiser. They take the place of good works, restraint, piety, awe, humility, and transcendence. On TV commercials, moral deficiencies as we customarily think of them do not really exist. A commercial for Alka-Seltzer, for example, does not teach you to avoid overeating. Gluttony is perfectly acceptable, maybe even desirable. The point of the commercial is that your gluttony is no problem. Alka-Seltzer will handle it. (Postman and Powers, p. 125)

These commercials emphasize removing the symptoms of moral deficiencies, rather than working on improving the virtues. Clearly this practice is objectionable from a virtue ethics standpoint. The problem with multivariate testing is similar. By implementing the results of the multivariate tests, a web developer could be understood as aiming to overcome virtues such as moderation, self-control, and self-discipline. Often the underlying question in a multivariate test is what one can do in order to encourage consumers to spend more money.

As with manipulative advertising, the virtue ethics argument against implementing results of multivariate tests is pretty straight forward

and shows the actions as clearly objectionable, because they hamper the development of virtues in the consumer. However, Virginia Held also offers another perspective from which to criticize the actions from a virtue ethics perspective: they undermine intellectual and artistic integrity (1984, 64-66). Although her writing focuses on advertisements, the criticisms may be ported over. The web developer and/or designer is in a real sense creating a work of art through the web page. Focusing on conversions and increased sales rather than the design itself is lacking in artistic integrity. Furthermore, intellectual integrity is also undermined.

The biggest problem with putting forth a virtue ethics critique such as this is the possibility of a response that implementing the results of multivariate tests actually highlight virtues. As previously mentioned, most of the literature on such tests frames them in a way that virtues such as altruism could easily be highlighted. A tendency could develop in which one views the situation as being merely relative: some people see such actions as manipulative while others see them as altruistic. Such a reading of the situation is likely, because one can certainly find examples of cases where multivariate testing is being used in a manner that is altruistic. The problem which may elude one who is not paying close attention to the issue is that not all uses of multivariate testing are equal.

Once again an issue of demarcation arises. Which uses of multivariate testing highlight positive virtues and which are problematic? If the discussion and reflection on such a question is left entirely up to those businesses doing the testing, recent history seems to suggest that all such testing will be championed as virtuous, but this is clearly not correct. To get a handle on the battle of language really going on, consider this report on the benefits of multivariate testing:

When it comes to segmenting, a large dating site found that those visiting its page during normal

business hours were receptive to a main photo depicting a happy couple, while those visiting in the evening were much more receptive to a photo of three provocative-looking blond women. The psychographic implications of that particular finding opened a treasure trove of marketable data for the company. (Quant is King, 2007)

Consider the language being used here. The visitors were 'more receptive' and the findings opened a 'treasure trove' of data. With this type of language it sounds as if the lives of everyone involved in the process are being improved. The company has access to important and valuable data and the visitors are being helped out by images with which they associate more pleasantly. But the alternative reading is that this valuable data is being used by companies to manipulate consumers. Segmentation of visitors to different times of the day allows for even greater manipulation. The "psychographic" implications on this reading are not a treasure trove of data, but rather a worryingly large amount of information that is being used to manipulate visitors into continuing to use the site and likely purchase a membership or view more advertising. Is the visitor actually benefiting? From a virtue ethics perspective that question is actually irrelevant. This type of manipulation implements data in a way that strives to weaken self-control and moderation from the visitors. Billing this practice as mutually beneficial is certainly not showcasing intellectual integrity.

Issues, Controversies, Problems

Perhaps the greatest challenge when considering A/B and multivariate tests is trying to get businesses to understand the subtle ethical considerations involved. First, most businesses are going to focus on the very pragmatic issue of increasing sales above and beyond all other considerations. However, even if a business is attempting to behave ethically, the ethical issues involved in multivariate testing can be easily written off in light of activi-

ties which appear much more blatantly unethical. Google, whose unofficial company motto is "do no evil," is the company that offers free multivariate testing for any website.

As demonstrated earlier through the words of Chuck Casto, Vice President of Corporate Communications for CSN Stores, one can easily summarize multivariate testing in a way that makes the actions appear to be altruistic rather than manipulative. Through this interpretation websites are simply trying to help customers achieve the actions they already want to perform. Because this is likely to be true in some cases, it is very difficult to determine where to draw the line: which tests lead to results that are actually helpful and which lead to the manipulation of customers. Currently there is no clear answer to this question. It is extremely unlikely that any company is going to completely eliminate multivariate testing from their web development strategy – the activity is actually becoming more popular than ever.

If any progress is going to be made in spreading the notion that this practice is unethical, clearer guidelines need to be developed which can differentiate between those uses which are ethically problematic and those which are not. A deontological system seems to be the best suited for developing such guidelines, but others may be possible. Deontology may be fruitful because a designer should be able to step back and examine the intentions of changes he or she is making to a particular page. Future work on this issue may be able to better explore the practical issue of ethical boundaries for the use of multivariate testing. One strategy that has been adopted widely in cases where testing on people occurs is informed consent. Currently, it is difficult or impossible for a typical consumer to tell if a website is conducting A/B or multivariate testing. It would at least be a step in the right direction if a website were to post a notice informing a visitor that the testing was occurring. This would alleviate some of the ethical problems, but others would remain: a

deontologist could still claim, for example, that the website was using the consumer merely as a means to increased profit.

Solutions and Recommendations

Multivariate testing seems more popular than focus groups, likely because it is both easier and cheaper to conduct as well as being more pragmatic. As with manipulative advertising, multivariate testing seems to be a practice that is clearly legal, but ethically questionable. Manipulative advertising has become so widespread that it is extremely unlikely there will ever be a reduction or elimination of such ads. Multivariate testing for websites is still a relatively new practice, so there may be a greater hope of intervening at this point in time.

If a professional code of conduct for multivariate testing could be developed, this may help stem the tide of its unethical use. Many companies offer multivariate testing as a third party service, so if they were to adopt a code of conduct, this would greatly reduce the number of companies participating in such a use, however, it would not prevent any particular company from undertaking their own internal multivariate testing.

FUTURE RESEARCH DIRECTIONS

This discussion would benefit greatly from empirical research into the question of why certain changes influence consumers to change their habits in the way that they do. The question of why may not be important for some ethical theories such as deontology, which is based on intentions, but those answers may be important for evaluation within other ethical frameworks.

Very little literature currently exists on the ethics of A/B or multivariate testing. However, the ethical issue is similar to that of manipulative advertising, therefore a broad understanding of the issues associated with manipulative advertising

would stimulate the understanding and discussion of the ethical issues that arise from the manipulation that occurs through multivariate testing.

CONCLUSION

Previously the discussion of multivariate testing has been left up to corporations and other businesses who charge to implement such tests. Little or no criticism of the practice has arisen. A/B and multivariate tests can be understood as problematic, at least in many uses, through three prominent ethical theories. However, if companies ever decided to include such tests with individualized data on consumers, the manipulation occurring would be even worse. Some segmenting is already done, as with the dating web site that draws a distinction between visitors during the day and visitors at night. But it is certainly possible to narrow those distinctions even further. Perhaps concerns about online privacy have attracted enough attention that this might not happen, but even without this connection, A/B and multivariate tests should be considered problematic.

Due to possible praiseworthy uses of such tests, some companies may not even realize that there are, or take time to reflect on, these ethical problems with such tests. Focusing on such an issue, even if it seems minor in light of other ethical concerns of e-commerce, may at least spread the notion that such tests may be used unethically so that those companies concerned with behaving ethically can stop to consider their own use of such tests.

It seems that there are clear cases of uses that are ethically praiseworthy and clear cases of uses that are ethically blameworthy, but in the middle will exist many uses which cannot be so easily categorized. Further effort at evaluating such middle cases through various ethical theories may be beneficial to all. At the very least, a better understanding of why such changes cause consumers

to increase purchases could be beneficial. Such answers could be incorporated into current media awareness campaigns, such as that by the Center for Media Literacy. Understanding the process and being aware of the causes and implications of A/B and multivariate testing would be a major step forward.

REFERENCES

- Arthur, C. (2009, July 8). *Google's Marissa Mayer on the importance of real-time search*. Retrieved September 7, 2009, from <http://www.guardian.co.uk/technology/2009/jul/08/google-search-marissa-mayer>
- Brohan, M. (2007). *Internet Retailer Survey: Form and function*. Retrieved September 7, 2009, from <http://www.internetretailer.com/article.asp?id=23262>
- COPPA - *Children's Online Privacy Protection*. (n.d.). Retrieved September 11, 2009, from <http://www.coppa.org/comply.htm>
- Crisp, R. (1987). Persuasive Advertising, Autonomy, and the Creation of Desire. *Journal of Business Ethics*, 6(5), 413-418. doi:10.1007/BF00382898
- Genge, W. (1985). Ads stimulate the economy. *Business and Society Review*, 1(55), 58-59.
- Google. (n.d.). *Google Analytics - Case Studies - American Cancer Society*. Retrieved June 13, 2009, from http://www.google.com/analytics/case_study_acs.html
- Google. (n.d.). *Google Analytics*. Retrieved September 7, 2009, from <http://www.google.com/analytics/index.html>
- Held, V. (1984). Advertising and program content. *Business & Professional Ethics Journal*, 3(3/4), 61-76.

- Internetretailer.com. (2008, May 29). Multivariate testing produces winning product pages for ShopNBC.com & InternetRetailer.com, *Daily News*. Retrieved May 28, 2009, from <http://www.internetretailer.com/dailyNews.asp?id=26545>
- Internetretailer.com. (2009, March 17). *Marketers plan to spend more online and aim to measure results better*. Retrieved September 7, 2009, from <http://www.internetretailer.com/dailyNews.asp?id=29781>
- Jhally, S. (Director). (1998). *Advertising and the End of the World* [Documentary]. USA: Media Education Foundation.
- Kant, I. (1993). *Grounding for the Metaphysics of Morals With on a Supposed Right to Lie Because of Philanthropic Concerns*. Indianapolis, IN: Hackett Pub Co Inc.
- Meyers, D. (n.d.). *The Secret to Happiness, YES! Magazine*. Retrieved June 9, 2009, from <http://www.yesmagazine.org/article.asp?ID=866>
- Mill, J. S. (2004). *Utilitarianism*. Boston: Public Domain Books. doi:10.1522/cla.mij.uti
- Montecino, V. (1996). *Copyright and the Internet*. Retrieved September 11, 2009, from <http://mason.gmu.edu/~montecin/copyright-internet.htm>
- Phillips, M. (1994). The Inconclusive Ethical Case Against Manipulative Advertising. *Business & Professional Ethics Journal*, 13(4), 31–64.
- Postman, N., & Powers, S. (1992). *How to Watch TV News*. Boston: Penguin (Non-Classics).
- Roche, M. (2004, July 26). *Scientific Web Site Optimization using AB Split Testing, Multi Variable Testing, and The Taguchi Method* | *WebProNews*. Retrieved June 12, 2009, from <http://www.webpronews.com/topnews/2004/07/26/scientific-web-site-optimization-using-ab-split-testing-multi-variable-testing-and-the-taguchi-method>
- Search Engine Marketing (SEM) Glossary*. (n.d.). Retrieved September 7, 2009, from <http://www.anvilmediainc.com/search-engine-marketing-glossary.html>
- sitespect.com (n.d.). *Online Businesses Doing More With Less In An Uncertain Economy Via Web Optimization Technology*. Retrieved May 20, 2009, from <http://www.sitespect.com/news-online-business-121608.shtml>
- Sneddon, A. (2001). Advertising and Deep Autonomy. *Journal of Business Ethics*, 33(1), 15–28. doi:10.1023/A:1011929725518
- Stoll, K. (2004, August 18). *Color your hunger*. Retrieved September 13, 2009, from http://findarticles.com/p/articles/mi_qn4179/is_20040818/ai_n11815804/
- Technology Weekly. (2007, May 30). *Quant is King*. Retrieved June 13, 2009, from <http://technologyweekly.mad.co.uk/Main/Home/Articlex/6f452b7cb4db47a8afc0ff9b8fb9752/Quant-is King.html>
- United States Access Board*. (n.d.). Retrieved September 11, 2009, from <http://www.access-board.gov/>
- Whitfield, T., & Wiltshire, T. (1990). Color psychology: a critical review. *Genetic, Social, and General Psychology Monographs*, 116(4), 385–411.