

AI and the Advent of the Cyborg Behavioral Scientist

Web Appendix

This Web Appendix is intended to illustrate some of the noteworthy results and observations arising from our process. We discuss most points in summary form, but refer interested readers to the following link, which houses the details of our interactions with the AI platforms including prompts and results from each stage of the process:
<https://github.com/g-tomains/Cyborg-Behavioral-Researcher.git>

Ideation Stage

In this stage, we wanted to get a sense of the relevance and quality of ideas that these AI models could put forth. We began by stating our evaluation criteria: importance in the literature, room for exploration, and many unanswered questions in the domain. We found that the models understood the scope of Consumer Behavior research fairly well, posing questions around emotion in decision making, social media, personalization and other relevant topics. Their initial responses were quite broad, to the point of being untestable without more specificity. That said, we were fairly satisfied with their ability to drill down to more refined ideas upon further prompting. We did not find any of the more specific ideas to be of exceptionally high quality, but thought many of them were passable enough that, with some expertise and attention, they could be developed into interesting projects. It was one of these ideas from ChatGPT that we ultimately pursued for this paper.

Illustrative initial responses from AI models:

Bard / Gemini - The role of emotions in consumer decision-making

Emotions play a major role in consumer decision-making, but there is still much to learn about how they work and how they can be influenced by marketers. For example, researchers could explore the following questions:

- How do different types of emotions (e.g., positive, negative, neutral) affect consumer decision-making?
- How can marketers use emotional appeals to influence consumer behavior?
- How do emotions interact with other factors, such as cognitive factors and social factors, to influence consumer decision-making?

Bing Chat / CoPilot - The effect of culture on consumer behavior: Culture is an important factor that influences consumer behavior, but there are still many unanswered questions about how culture affects consumer decision-making. This topic could explore how cultural differences affect consumer behavior and how companies can adapt their marketing strategies to different cultures.

Illustrative responses from AI models after a more specific request:

Bard / Gemini - Consumers will be more likely to prioritize ethics over price for products that they perceive to be essential or high-impact. For example, consumers may be more willing to pay a premium for ethical food or clothing than for ethical electronics or luxury goods.

ChatGPT 4 - ****Ethical Fatigue****: There will be a segment of consumers who experience 'ethical fatigue' and become skeptical or indifferent to ethical branding due to overexposure to marketing messages about ethics.

Literature Review Stage

We found the literature review stage to be a major weak point for the AI models. Firstly, the hallucinations they made, i.e., inventing research that does not exist, made them highly unreliable. While a researcher could take a “trust, but verify” approach to this stage, the effort it would take to do so very much undermines the time-saving value that these models are proposed to offer. Furthermore, these models often referred us to a database or search strategy, rather than offering specific suggestions. Moreover, the legitimate research that these models did recommend was rarely from relevant, top-tier journals, again making this a poor utility to researchers. Last and most crucial, the relevance of the research these models did raise was frequently only relevant at a superficial level.

Illustrative responses from AI models:

Bard / Gemini - I can't conduct a full academic literature search myself, but I can definitely help you get started. Here are some ways to find relevant research on "ethical fatigue" in the context of consumer behavior:

Search Engines: Use Google Scholar with terms like "ethical fatigue consumer behavior," "ethical marketing overexposure," or "consumer desensitization to ethical claims."

ChatGPT 4 - ****Ethical Consumer Behavior****: A study by Newcastle University Business School focused on the motivations behind purchasing ethically produced goods, introducing concepts like "conspicuous ethics" and an "ethical index" ([Newcastle University Business School](https://www.ncl.ac.uk/)). This research primarily concentrates on the positive aspects of ethical consumption, such as consumer preferences for ethically made goods and the willingness to pay more for such products.

(Note: The link ChatGPT 4 provided was simply a link to Newcastle University's homepage.)

Research Design Stage

Methods

We found the models to be particularly impressive at developing methods. They exhibited a reasonable degree of knowledge around what makes a prototypical behavioral experiment. Moreover, they adequately adapted this knowledge to the needs of the hypothesis we wished to test. We specified some of the requisite details and context, but left the rest to the AIs. We found the methods the models came up with to be fairly close to what we personally would have developed to test this hypothesis although all responses required some degree of adaptation. We found the stimuli they generated likewise generally very useful.

Our prompt for a design: In a previous conversation, you gave me the following research idea. Please give me 5 between-subjects, 2-cell online behavioral lab experiments I can run to test this main prediction: **Ethical Fatigue**: There will be a segment of consumers who experience 'ethical fatigue' and become skeptical or indifferent to ethical branding due to overexposure to marketing messages about ethics.

Illustrative responses to our prompt:

Bard / Gemini - **Experiment 1: Ethical Message Overload**

Hypothesis: Consumers exposed to a high volume of ethical marketing messages will show increased skepticism and indifference towards ethical branding compared to those exposed to a low volume of ethical marketing messages.

Design:

- Participants will be randomly assigned to either a high-ethical-message condition or a low-ethical-message condition.
- Participants in the high-ethical-message condition will be exposed to a series of advertisements and social media posts that emphasize the ethical practices of brands.
- Participants in the low-ethical-message condition will be exposed to a series of advertisements and social media posts that do not emphasize the ethical practices of brands.
- After exposure to the marketing messages, participants will complete a survey measuring their attitudes towards ethical branding.

Bing Chat / CoPilot - Experiment 1: Manipulate the frequency of exposure to ethical branding messages by randomly assigning participants to either a high-exposure condition (where they see 10 ads from different brands that emphasize their ethical values) or a low-exposure

condition (where they see 2 ads from different brands that emphasize their ethical values). Then, measure their skepticism and indifference towards ethical branding using a scale.

(Note: The last sentence contained an embedded link which goes to a webpage about ethical marketing, but does not appear to contain a scale.)

Stimulus Design

As with designing general methodologies, we found the AI models highly capable at designing usable stimuli. They demonstrated significant utility as tools for this purpose, especially for experiments that involve variegated stimuli, such as this one.

Illustrative responses:

ChatGPT 4 (generating statements for non-ethical branding stimuli) - Sure, here are brand statements for four different companies, focusing on their product attributes without referencing ethics or including Nike:

1. ****Canon****: Canon's cameras offer unrivaled precision and clarity, capturing moments with detail that feels almost more real than life. Whether you're a professional photographer or an enthusiastic amateur, Canon's range of DSLRs and mirrorless cameras are designed to inspire creativity and push the boundaries of photographic excellence.
2. ****Bose****: Bose delivers an auditory experience that transcends the ordinary, enveloping you in a landscape of pristine sound. With cutting-edge technology, Bose headphones and speakers blend innovation and artistry to produce unparalleled sound fidelity, whether you're at home, in the office, or on the move.
3. ****Lego****: Lego sets unlock a world of imagination, offering endless possibilities for creation, from the simplest structures to intricate masterpieces. Known for their durability and precision interlocking, Lego bricks empower builders of all ages to explore their creativity and bring their visionary ideas to life.
4. ****Dyson****: Dyson's line of vacuums and air purifiers epitomize the fusion of design and functionality, offering powerful, efficient cleaning with a sleek, modern aesthetic. Engineered for performance and ease of use, Dyson products are a testament to the brand's commitment to improving everyday life through innovative technology.

Bard / Gemini (suggestion for dependent measures) -

Brand Perception Questionnaire

Please rate your perceptions of the target brand using the following scale:

1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree

- The target brand is ethical.
- The target brand cares about the environment.
- The target brand is socially responsible.
- I would trust the target brand to do what is right.
- I would recommend the target brand to others.

Additional Measures

- Purchase intent: Would you be more likely to purchase a product from the target brand after seeing their ethical messaging?
- Brand loyalty: How loyal are you to the target brand?
- Brand favorability: How favorable is your overall impression of the target brand?

Coding

In this task, we provided the AIs with the design and stimuli as previously determined and asked them to provide us with a fully functional Qualtrics QSF file with which to run the study.

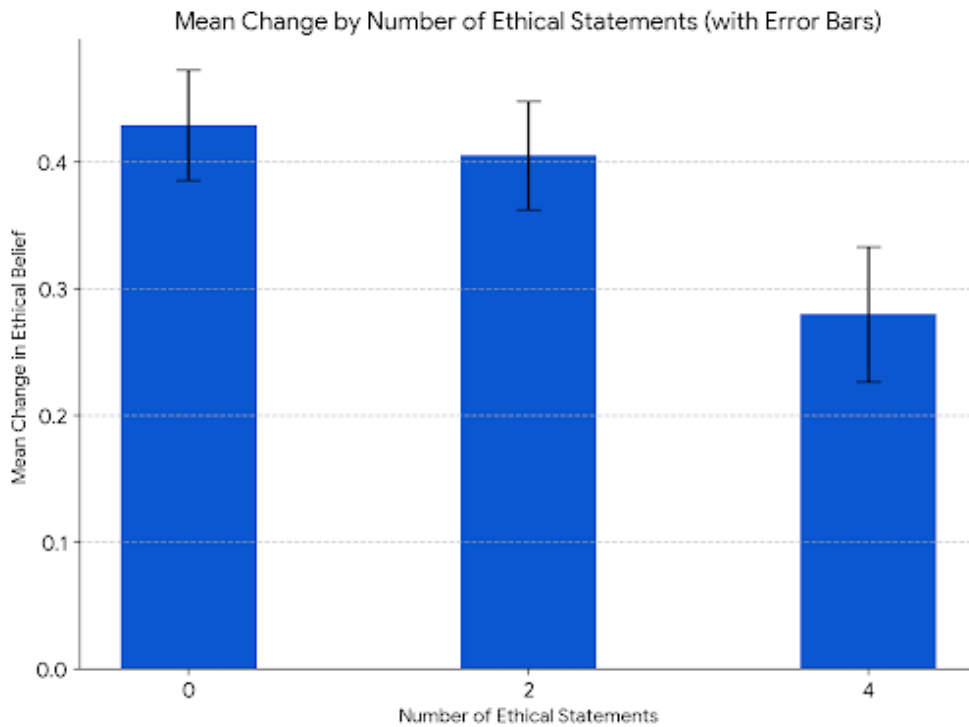
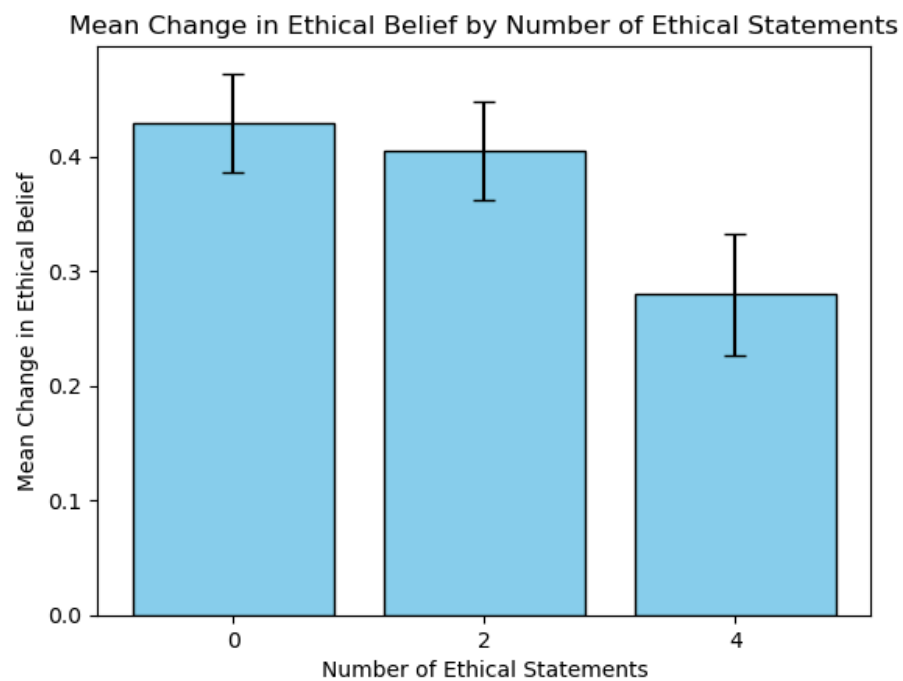
Bard/Gemini refused to produce a QSF file outright citing unspecified security restrictions, while the other two AIs attempted to provide QSF files but ultimately failed. In lieu of the requested file, the AIs instead provided lengthy walkthroughs for how to set up such a file on our own. This seems like a reasonable substitute on the face of things, but a closer examination of these proposed procedures shows that they are little more than restatements of the design. They have little in the way of specific suggestions regarding *how* to achieve these designs in Qualtrics. We attempted to exploit coding ability further with ChatGPT 4 by asking it to create a QSF file for a very simple survey. It was able to produce this file, but the file was unable to be parsed by Qualtrics. Closer inspection of the file indicated that it was missing many tags and blocks that would be necessary for a QSF file. We repeatedly tried to have ChatGPT 4 reformat the file, but were never able to obtain anything close to a usable QSF file.

Results Stage

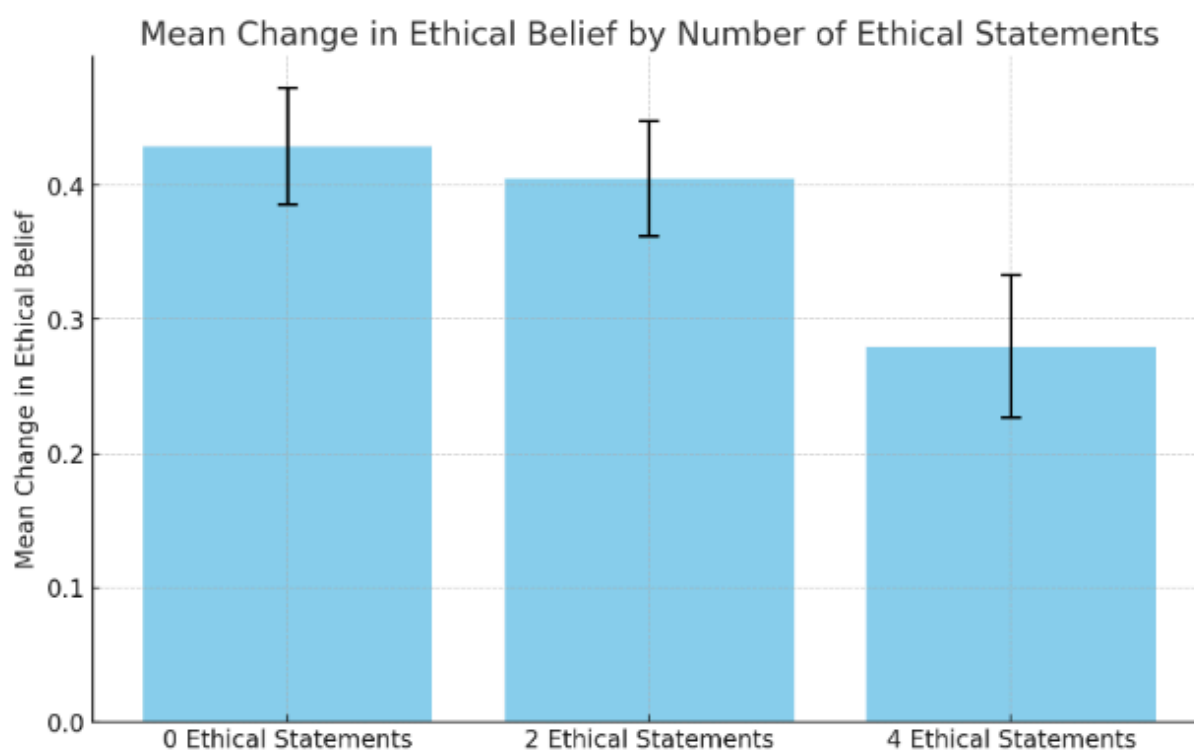
In this stage, we provided the AIs with a CSV file of the results of Study 1 exported directly from Qualtrics and asked the AIs to analyze the data to produce appropriate statistical tests and charts. We found it impossible to upload CSV files to Bard/Gemini at that time making further analysis impossible. Bing/CoPilot allowed us to upload the CSV file, but was unable to process the file. We repeated the same prompt adding that if it was unable to analyze the data, that it should provide a python script that could. It provided a reasonable script for the basic analysis. In contrast, ChatGPT 4 was able to perform the analysis itself, but when we asked for specific reportable statistics, we found that it had inexplicably used only a subset of the data for the analysis.

Notably, all three AI's recognized that ANOVA was the appropriate default model for performing the analysis. We were curious about whether the analyses suggested/performed considered potential violations of the model, so we asked for follow-up analyses from ChatGPT 4. It recognized a subset of the applicable assumptions, but was unable to perform those tests for us. It also struggled to provide us with appropriate test statistics for reporting our results, which also limited our ability to detect its analytic errors.

Because the AIs provided erroneous analyses (if at all), we used other tools to perform the analyses ourselves for the results section of the embedded paper. But we still wanted to present visualizations of the effects, so we used the AIs for this narrow task, providing them with the relevant quantities to plot. All three AIs provided reasonable visualizations, although we mistakenly started using ChatGPT 4o, which would not provide error bars. But when we switched models to ChatGPT 4, it provided an appropriate plot.

Example chart outputs:**Bard / Gemini -****Bing Chat / CoPilot -**

ChatGPT 4 -



Extension Stage

Here we again saw the models' abilities to design an experiment well around a given idea. However, the ideas for extensions were themselves of minimal value. In particular, we asked the AI for ideas on how to explore the scope and mechanism of the effect more. Most of the AIs offered no intuitive advancement of the initial study, primarily recharacterizing the main effect. Moreover, the suggestion that we ultimately did attempt, as we expected, ultimately did not yield a significant interaction.

Another interesting impediment was that the AIs appeared not to understand what was meant by a "moderation" study or effect. We had to spend substantial effort specifying the concept of an interaction and that it needed to be tested by crossing our existing manipulation with another factor of theoretical interest in a factorial design.

Illustrative moderation ideas:

Bing Chat / CoPilot -

Cognitive Dissonance:

- Exposure to multiple ethical brand statements could have triggered cognitive dissonance. Participants might have experienced conflicting thoughts about Nike's ethics.
- When faced with too much pro-ethical information, individuals may unconsciously seek balance by downplaying the impact of any single statement (including Nike's).

Bard / Gemini -

Sample Size and Demographics: The sample size and demographics of participants can influence the results. Consider exploring these factors for potential variations.

(Note: Because the responses with each AI at this stage varied, we found that our prompt sequence here also varied more than in prior stages. We ultimately settled on an extension idea provided by ChatGPT of crossing the existing ethical statement frequency manipulation from study 1 with a message framing manipulation. But, as with the other AIs, it took considerable effort to get there. We began by asking for a moderation study, and it supplied us with a four-cell confounded design):

****Experimental Conditions**:**

1. ****Low Exposure (Control Group)**:** Participants are exposed to 0 ethical statements from other brands.
2. ****Moderate Exposure**:** Participants are exposed to 2 ethical statements from other brands.
3. ****High Exposure (Saturation Group)**:** Participants are exposed to 4 ethical statements from other brands, as in the original study.
4. ****Varied Content High Exposure**:** Participants are exposed to 4 statements from other brands, but only 2 are about ethicality, and 2 are about unrelated aspects (e.g., innovation, performance).

We continued by explaining that we were looking for a design that crossed two factors ideally in a 2x2 design. It responded with a study using an individual difference variable, ethical sensitivity. It provided some useful approaches for measuring ethical sensitivity, but we ultimately asked it to use a factor that we could manipulate experimentally, to which it responded:

Second Factor: Message Framing

****Definition**:** Message Framing refers to the way information is presented to participants, particularly in terms of its positive (gain) or negative (loss) implications. In the context of ethical brand statements, positive framing might emphasize the benefits of ethical practices (e.g., environmental preservation, social welfare improvements), while negative framing might highlight the consequences of unethical practices (e.g., harm to the environment, exploitation of workers).

2x2 Factorial Design

****Experimental Conditions**:**

1. ****Prior Ethical Statements Exposure**:** 0 (control) vs. 4 (high exposure)
2. ****Message Framing**:** Positive (gain) vs. Negative (loss)

Furthermore, it predicted two main effects and the following interaction effect:

- H3: There will be a significant interaction effect, where the impact of exposure to multiple ethical statements is more pronounced under negative framing compared to positive framing.

Apart from the vague operationalization of framing, This seemed like a feasible if not especially plausible direction for extension. Note, however, when asked to explain why it predicted H3, it responded with the vacuous:

****Rationale**:**

This design allows for examining how the presentation (framing) of ethical information interacts with the quantity of ethical exposure to influence perceptions of Nike's ethicality. It provides insights into not just whether saturation occurs, but how the manner in which ethical content is communicated can modulate this effect. This can shed light on strategic communication approaches for brands aiming to highlight their ethical practices effectively.

Manuscript Production Stage

Producing the manuscript required uploading documents, so the AI would have the relevant information to construct the manuscript (e.g., the main idea, empirics, etc.). We found, however, that only ChatGPT 4 and Gemini could receive documents this way. As such, we only utilized these models for manuscript production. As we found the output

from ChatGPT 4 to be superior, we opted to include that in the manuscript. Also of note, we tried giving both ChatGPT 4 and Gemini all of the necessary information to produce a manuscript. ChatGPT 4 and Gemini were unable to process a document of this size, so we instead performed this for separate sections of the document.