Neuromarketing Beyond the Posttest: fMRI Can Predict the Commercial Effectiveness of Storyboards Before the TV Commercial Is Shot

H. Steven Scholte, Andries van der Leij, and Victor A. F. Lamme

1 Introduction

What makes for a good TV commercial (TVC)? The question itself can already be understood in very different ways. What is "good"? A commercial that wins prizes for creativity or originality? Or does a good commercial simply increase sales or brand value? Obviously, it is best when the two go hand in hand. However, while advertising agencies often know very well what makes for an original or creative design, there is less agreement on what the ingredients are of a TV commercial that pays off in the more literal sense. Given the high costs of advertising, it is paramount to determine this as early as possible. Therefore, the pretesting of TV commercials is common practice, given that airing typically is most costly (Fig. 1). Also, this allows for the allocation of marketing budget to the most effective ways of influencing consumers' decisions, i.e., the commercials that can be expected to "work" best. Even so, it remains problematic having to conclude that a TV commercial is better not aired (or with limited budget) after the costs of production have been made. Some TV commercials take hundreds of thousands to make. It is possible to alter a TV commercial after production, by reediting, using different music, altering voice-overs, or changing the pack-shot. If guided by proper (neuro-)marketing research and expertise, this approach can sometimes improve the commercial considerably. But even more helpful would be a tool that allows for the testing of effectiveness *prior* to production.

H. S. Scholte (✉) · V. A. F. Lamme
Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

Neurensics BV, Amsterdam, The Netherlands
e-mail: h.s.scholte@uva.nl

A. van der Leij
Neurensics BV, Amsterdam, The Netherlands

B. B. Briesemeister, W. K. Selmer (eds.), *Neuromarketing in Business*, Management for Professionals, https://doi.org/10.1007/978-3-658-35185-4_4

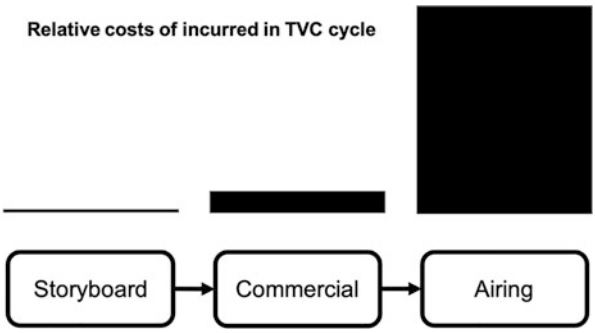


Fig. 1 The life cycle of a TVC, from storyboard to airing. Costs of different phases rise exponentially

In this paper we will demonstrate that a) it is possible to predict the success of a TVC on the basis of BOLD-MRI recordings (a branch of neuromarketing research popularly known as "brain scans"), and second, given this approach, it is possible to predict this success already on the basis of the storyboard.

2 TV Commercial Effectiveness Predicted by Functional MRI

Functional MRI-based neuromarketing techniques have proven a particularly useful tool, given its capability to predict effectiveness in a highly reliable way. For example, it has been shown that the ability of anti-smoking campaigns to induce a change in smoking behavior is better predicted by MRI recordings of brain activity, than by the opinions of smokers or marketing experts (Falk et al. 2012; Sherman et al. 2018; Gearhardt et al. 2020). Given these results, neuromarketing is regarded by one part of the marketing professionals as holding great promise for the future of design and marketing. Others are more skeptical or think that the techniques and possibilities are still in their infancy (Hubert and Kenning 2008; Ariely and Berns 2010; Berkman and Falk 2013).

Venkatraman et al. (2015) performed a large comparative study of several neuromarketing techniques to study their relation with traditional research methods in response to a set of 30-second ads. Importantly, they also studied to what extent the measurements were able to predict ad elasticity: the percentage change in sales due to the change in ad expenditure or gross rating points. Functional MRI (fMRI) was shown to be the only technique to explain significant added variance in ad elasticity beyond traditional measures. While traditional measures (self-reported liking, relevance, purchase intent) explain about 0.65 (R^2) of ad elasticity, explained variance goes up to 0.87 (R^2, $p = 0.034$) when combined with fMRI, in this case activation of the amygdala, ventral striatum, dorsolateral prefrontal cortex, and ventromedial prefrontal cortex. Other neurophysiological techniques like eye-tracking and EEG and physiological responses like SCR and heart rate did not explain added variance beyond the self-report measures. This landmark paper proved the added value of BOLD-MRI in determining the effectiveness of **produced** commercials.

3 The Elusive Ingredient of Effectiveness

In the Netherlands, prizes are awarded for TV commercials on the basis of effectiveness, likability, and annoyance. To win an effectiveness award (the "Effie"), advertising agencies send in their TV commercials together with evidence substantiating the claim that the commercial led to an increase in sales, recognition, or valuation of the brand or product the commercial was for. A professional jury looks at each case and decides who wins. Likability awards (the "Gouden Loeki") are voted for by the general public, and winners are announced in a festive program on national TV. Awards for the most annoying TV commercial (the "Loden Leeuw") are also voted for by the general public and are announced in a consumer interest TV program.

Although each of these awards is based on a variety of aspects, they also form distinct categories. Gouden Loekis are typically won by TV commercials with high gross rating points (GRP) that are funny, cute, or otherwise likeable. Loden Leeuwen typically are high GRP ads that annoy, for all sorts of reasons. The Effie winners are much more difficult to describe as a category. Their distinctive feature obviously is their capability of increasing sales or brand loyalty, but it is not so easy to get a grip on how this is achieved. They are often not particularly funny, and sometimes they are esthetically pleasing or technically well made, but Effie awards have been won by very non-outspoken TV commercials. So what is their secret? To study this, we have investigated the neural signatures of the three types of TV commercials using BOLD-MRI.

4 Effective TV Commercials Have a Unique "Neural Signature"

For this study, we combined brain activity recorded during 6 multiclient and internal research studies in each of which we recorded from 16 to 27 subjects. Subjects were exposed to 8–16 TV commercials, in addition to 4–6 benchmark commercials (Effies, Gouden Loekis, Loden Leeuwen). Furthermore, we performed 1 study (16 subjects) in which we only used Effies, Loden Leeuwen, and Gouden Loekis (8 each). Results from the latter study were used for the whole brain analysis presented in Fig. 2. In total (obtained over 119 subjects), we thus obtained 17 recordings of Effies (many of the same ad but in different pools of subjects, 8 unique ads), 20 "Gouden Loeki" recordings (8 unique ads), and 20 "Loden Leeuwen" recordings (9 unique ads). We analyzed BOLD-MRI responses to these TVCs on different aspects.

A first thing to note is that TVCs are very effective in activating the entire brain (see Fig. 2, top). Typical laboratory studies, using artificial and less engaging stimuli, show much more limited brain activation. Figure 2 (bottom) shows for where the activations differ between effective, likable, and annoying TVCs. Effective commercials stand out on the basis of their ability to activate brain regions such as the orbitofrontal cortex, and subcortical areas known to be involved on the valuation

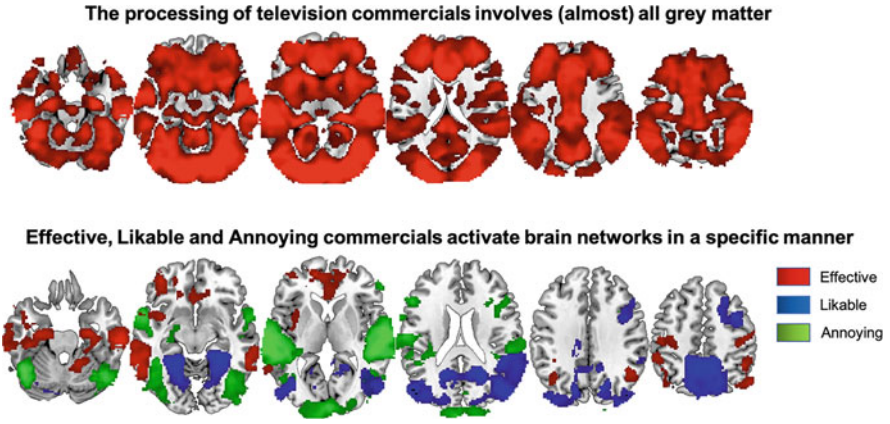


Fig. 2 The neural signatures of effective (blue), likable (red), and annoying (green) TV commercials. Note how effective commercials (blue) activate prefrontal and deeper subcortical regions associated with emotional valuations such as expectation, reward, and desire. Likable commercials (red) activate visual association areas and parietal regions involved in attention. Annoying commercials (green) are characterized by a strong activation of early visual and auditory areas, indicating that they almost literally strain the eyes and ears (whole-brain analysis of 8 TVCs of each class, 16 subjects, recorded on a 3 T Achieva TX, 32 channel head coil, Philips)

of reward, expectation, and desire, and found to predict consumer conversion (Plassmann et al. 2008; Berns and Moore 2012; Falk et al. 2012; Gearhardt et al. 2020).

To quantify the differences between Effies and other award winning commercials, we analyzed the full set of 57 benchmark recordings using our standardized brain rating procedure (Fig. 3). In this analysis, we first map, in each subject, the regions of the brain involved in 13 different brain networks related to emotions, processing, and valuations using either standardized stimuli or proprietary stimuli and tasks developed in our lab. The possibility of a reverse inference of emotions from brain activity was verified using classification and correlation analyses. A description of these 13 networks is given in Fig. 4.

We then standardize the activations of these emotion related networks to obtain a "neuro-emotional signature" for each recorded TVC. Figure 3 shows the average neuro-emotional signatures of the three TVC types. Effies stand out from the other types of award winning TV commercials because they evoke higher levels of positive emotions like desire, lust, expectation, and trust while at the same time evoking lower levels of negative emotions like fear, anger, disgust, and danger.

What do we learn from this? Most remarkable is that effective commercials do not need to draw a lot of attention or personal interest. That explains why they are often quite inconspicuous. Negative emotions are typically avoided. Eliciting desire and expectation seems most important, together with evoking a sense of trust. In our sample, we have seen these emotions, for example, being brought about by portraying craftsmen or people in a family setting and by using easy listening-type

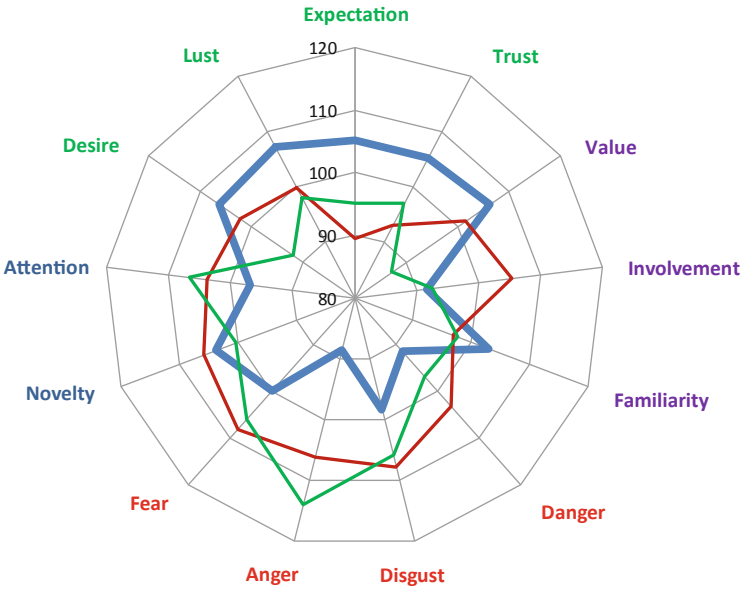


Fig. 3 Brain rating of effective (blue), popular (red), and annoying (green) TV commercials (weighted average of 17–20 TVC recordings of each class). See Fig. 4 for a description of the neural networks used for the classification

Networks	Positive emotions
Desire	Predicts levels of desire. Consists of cortical and subcortical areas.
Lust	Predicts levels of lust. Consists of cortical and subcortical areas.
Expectation	Predicts levels of expectation. Consists of cortical and subcortical areas.
Trust	Pedicts levels of trust. Contains, and links, hot and cold prefrontal areas.
	Affective processes
Value	Predicts levels of value. Includes the nucleus accumbens.
Involvement	Predicts levels of involvement. Includes so called 'resting state' areas.
Familiarity	Predicts level of novelty. Includes parts of the entorhinal cortex
	Negative emotions
Fear	Predicts levels of fear. Includes the amygdala.
Anger	Predicts level of anger. Similar to fear but with an action component
Disgust	Predicts disgust. Includes part of the insula.
Danger	Predicts ganger. Similar to fear but with a motor component.
	Impact
Attention	Predicts level of attention.
Novelty	Predicts level of novelty. Includes parts of the ventral visual cortex

Fig. 4 Short description of 13 networks related to emotion

music. Filming is generally of high quality and the cutting is of medium to slow pace, all contributing to a composed rather than exhilarating quality.

5 Can Effectiveness Be Predicted Using TVCs?

Now that the neural and emotional signatures of effective TV commercials are known, can this be used to predict effectiveness of a TVC in advance? We analyzed whether pretesting for effectiveness is possible. For each class of commercials, we constructed an average (template), leaving one of the commercials out of this average. The pattern of neural activity of the left out commercial was then correlated with each of the three templates. This was repeated for each commercial (Fig. 5). Correlation values were fed into a classification algorithm to categorize each commercial as either effective, likable, or annoying (leave-one-out classification, e.g., Ponseti et al. 2012).

Categorization was particularly successful for the Effie commercials: 15 of the 17 Effie commercials were indeed classified as being effective (88% hit rate), while 35 of the 40 non-Effie commercials were classified as not effective (87% correct rejections). This included correlations between tests of the same commercial but recorded in different pools of subjects (usually ~22 "average" consumers). Such test-retest classifications are typically identical (92%). But even if these commercials were left out, 6 of the 7 Effies (86%) and 10 of the 13 non-Effies (77%) were still correctly classified.

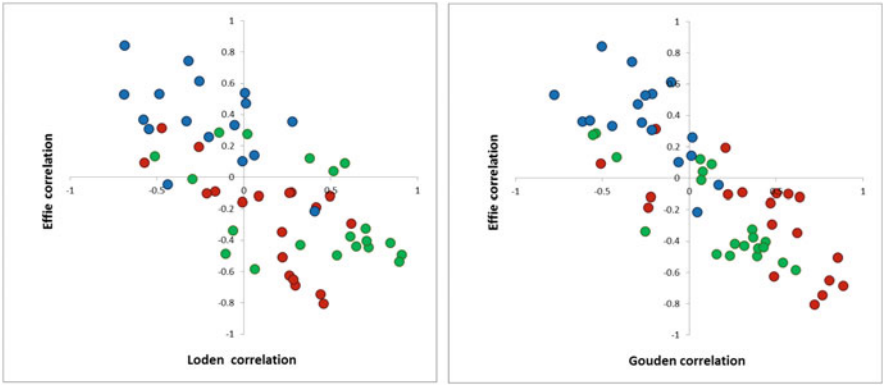


Fig. 5 Leave-one-out correlations for Effies (blue), Gouden Loekis (red), and Loden Leeuwen (green) show that effective commercials (blue circles) stand out as a distinct category from noneffective commercials (red and green). Effies (blue) have high internal correlation and are negatively correlated with the other two types of award winning commercials (red and green). Note that the graph depicts both test-retest correlations of the same commercials tested in different groups of subjects and correlations between different commercials of a particular category, tested in either the same or different groups of subjects

In sum, we show that effective TV commercials evoke a specific neural signature and stand out as a distinct category when rated via an fMRI-based neural analysis. This allows for the successful pretesting of the effectiveness of a TV commercial. Two conclusions can be drawn that are of particular importance to the neuromarketing field. The 92% correct repeat classification tells us that recording from ~22 subjects yields sufficiently stable results to reliably classify a TV commercial. That classification of different Effie commercials is almost as good (82%) tells us that this approach is both reliable and valid.

6 Reverse Engineering of Storyboards from TV Commercials

We have shown it is possible to successfully predict the effectiveness of a TVC with BOLD-MRI. But if we consider Fig. 1, it would be even better to predict the success of a campaign on the basis of a storyboard. A storyboard is created before the actual recording of the TVC and consists of a sequence of (animated) images, accompanied by a brief story explaining the narrative and general idea or concept. We tested to what extent these kinds of storyboards could predict the impact on the brain – and hence the effectiveness – of the actual full-production TV commercial.

Twelve actual TV commercials were reverse engineered to storyboards in a systematic fashion: key shots of the TV commercial were transformed into hand-made drawings, typically ten or so per commercial. These were accompanied by either a) a voice-over, explaining the action and reading the final pack-shot message, or b) the original soundtrack of the TVC. Drawings and sound were compiled into a slideshow, of equal length as the original commercial. Subjects, while lying in the MRI scanner, first viewed the 2x12 storyboard versions, followed by the 12 actual TV commercials. Note that subjects did not know these commercials, as they were foreign (mostly UK and US) commercials not aired in the Netherlands.

We evaluated the TVCs with the above method that is capable of predicting TVC success (i.e., using the 13 networks from Fig. 4). We correlated the neuro-emotional values obtained for storyboards with those obtained for the actual TV commercial, to test to what extent storyboards have any predictive value for the neural impact of TV commercials.

The example shown in Fig. 6 illustrates the principle of analysis. Both the TV commercials evoke fairly similar (but not identical) mapper values. Note, for example, how in both negative emotions are generally larger than positive emotions. This similarity can be expressed in the correlation value r between the two vectorized mapper scores, which in this case is 0.76. A value of 1.0 would indicate perfect similarity, while a value of −1.0 perfect dissimilarity (so that values are their exact opposite). This high similarity in response amplitude between storyboard and TVC is only obtained after a, within neural network, normalization procedure. Before this procedure, the responses to the storyboard are (as expected) much less intense than the responses to the TVCs.

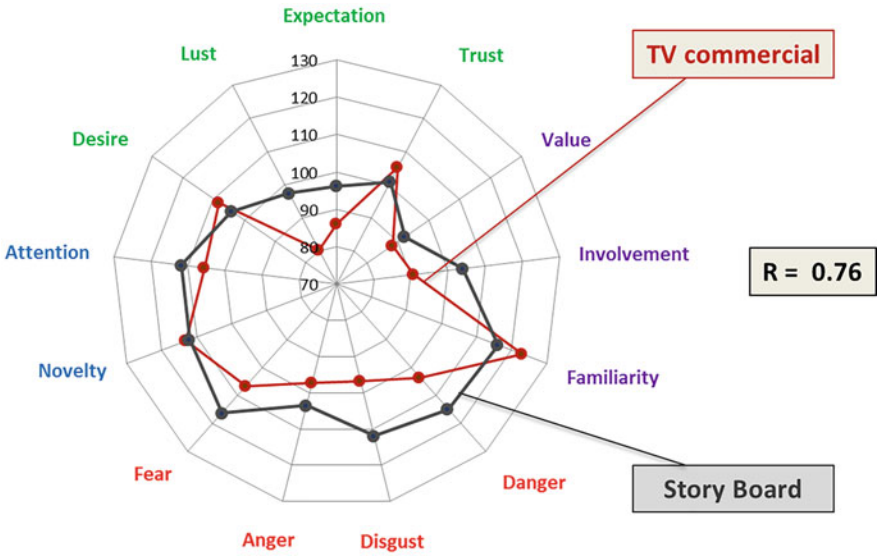


Fig. 6 Plot of 13 neural networks for an example TV commercial (red) and its storyboard version (gray). Values are not identical but fairly similar (note how in both negative emotions are higher that positive emotions) resulting in a correlation value (r) of 0.76

7 TV Commercials Are Predicted from Their Storyboards

In the sample of 12 TV commercials and storyboards we tested, we found cases of fairly high correlation and cases of nearly zero correlation, but never negative correlations (see Fig. 7). We included in our sample TVCs with (a) no clear auditory or visual narrative, (b) a clear auditory narrative with more "evocative" images, and (c) a clear auditory and visual narrative. The results show that our procedure failed for the first type of TVC (no clear auditory and visual narrative, commercials 1 and 6 in Fig. 7). The other two types of commercials showed high correlations between TVC and storyboard version (MSB), for both the storyboards with the original soundtrack and the narrative description (and no significant difference between them). On average the correlation value is 0.62. Statistically, this deviates strongly from change ($p = 0.00002$).

These results show that the relatively simple images and narrative of storyboards have a reliable predictive value for the effectiveness of the final TVC.

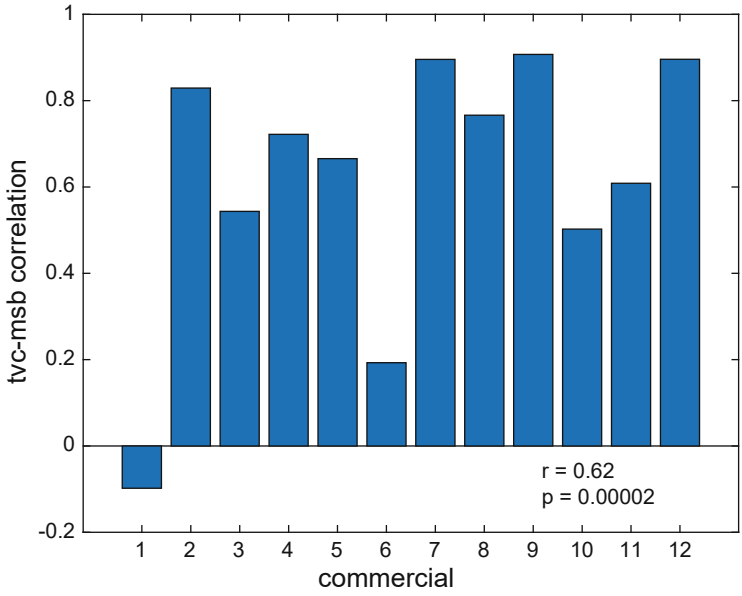


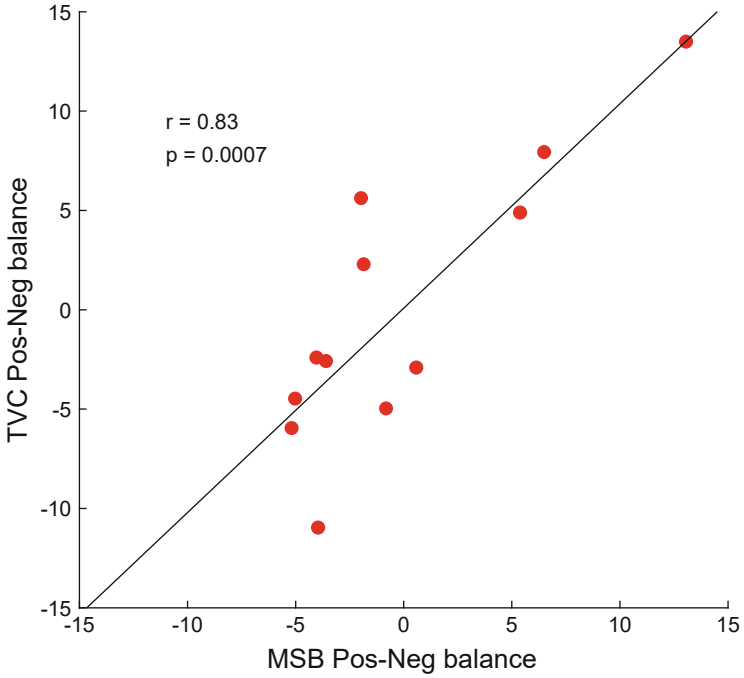
Fig. 7 Correlation values between the spider plots of 12 TV commercials and their corresponding storyboards. For the TVCs with a strong visual narrative, the correlation with the original track is shown. For the other TVCs, the correlation with the description is shown

8 Nearly Perfect Prediction in the Dimension That Matters: Pos-Neg Balance

Earlier research has shown that the most important parameter for the effectiveness of a TV commercial is the *balance* between positive and negative emotions (Lamme and Scholte 2013). Much more than the absolute values of these emotions, this balance predicts effectiveness: will consumers buy the product (or "buy" the message) and show behavioral conversion?

Why is that the case? Positive emotions can be high, but if these are offset by equally high negative emotions, the net result is a message that leaves mixed feelings. Therefore, consumer behavior is best predicted by the *net balance* of these two opposing forces (positive main axis value – negative main axis value).

We find that storyboards are very well suited to evaluate the future pos-neg balance of a TV commercial. The correlation value increases to 0.83 ($p = 0.0007$) when storyboards are compared to TV commercials along this dimension.



This has very important implications. The future effectiveness of a TV commercial (indexed by its pos-neg balance) turns out to be the value that is best predicted by its storyboard. This makes storyboard testing a most useful tool if TV commercials are targeted at real conversion of consumers. It is the power to convince, to induce a change in behavior, and to move consumers toward shops that is best evaluated using fMRI-based storyboard testing.

This adds to the growing evidence that the real power of fMRI-based neuromarketing lies indeed in its capability to predict consumer *behavior*, instead of consumers' opinions (Falk et al. 2012; Berns and Moore 2012).

9 Afterthoughts: Toward a New Way of TV Commercial Production

To return to the first figure of this paper, there is an exponential cost increase in using TVCs progressing from storyboard to campaign. In this paper, we travel from effective commercials in the market to predicting this success on the basis of the brain response to these TVCs. Finally, we close this loop by showing that the storyboards can also be used to evaluate if a TVC is going to be successful. This means that the tool of fMRI-based storyboard testing is well suited to compare different TV ad proposals at the level of storyboards when a decision needs to be made about which of these proposals will most likely result in consumer conversion.

This may require a change from common practice, where storyboards are compiled relatively late in the decision process: typically to brief production after a concept has been selected and fleshed out.

We propose that storyboards should be used relatively early. Already in the conceptual stage, when several creative directions are discussed, storyboards should be made for all directions under consideration. The storyboards can be tested against in a single fMRI run in a sample representative of the target group. The predicted effectiveness of the creative directions can help support decision-makers in choosing which direction to pursue. By doing this, fMRI storyboard pretesting has the maximum expected return on investment.

For the first time, a tool is available that, when properly used, can change the industry of TV commercial production. No longer do we need to wait for a commercial to work. We can already know it in advance.

References

Ariely, D., & Berns, G. S. (2010). Neuromarketing: The hope and hype of neuroimaging in business. *Nature Reviews Neuroscience, 11*, 284–292.

Berns, G. S., & Moore, S. E. (2012). A neural predictor of cultural popularity. *Journal of Consumer Psychology, 22*, 154–160.

Berkman, E. T., & Falk, E. B. (2013). Beyond brain mapping. Using neural measures to predict real-world outcomes. *Current Directions in Psychological Science, 22*, 45–50.

Falk, E. B., Berkman, E. T., & Lieberman, M. D. (2012). From neural responses to population behavior: Neural focus group predicts population-level media effects. *Psychological Science, 23* (5), 439–445.

Gearhardt, A. N., Yokum, S., Harris, J. L., Epstein, L. H., & Lumeng, J. C. (2020). Neural response to fast food commercials in adolescents predicts intake. *The American Journal of Clinical Nutrition, 111*(3), 493–502.

Hubert, M., & Kenning, P. (2008). A current overview of consumer neuroscience. *Journal of Consumer Behavior, 7*, 272–292.

Lamme, V. A. F., & Scholte, H. S. (2013). TV commercial effectiveness predicted by functional MRI. *Neuromarketing Science and Business Association Magazine*. https://silo.tips/download/tv-commercial-effectiveness-predicted-by-functional-mri.

Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences USA, 105*, 1050–1054.

Ponseti, J., Granert, O., Jansen, O., Wolff, S., Beier, K., Neutze, J., Deuschl, G., Mehdorn, H., Siebner, H., & Bosinski, H. (2012). Assessment of pedophilia using hemodynamic brain response to sexual stimuli. *Archives of General Psychiatry, 69*, 187–194.

Sherman, L. E., Hernandez, L. M., Greenfield, P. M., & Dapretto, M. (2018). What the brain 'likes': Neural correlates of providing feedback on social media. *Social Cognitive and Affective Neuroscience, 13*(7), 699–707.

Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., Hershfield, H. E., Ishihara, M., & Winer, R. S. (2015). Predicting advertising success beyond traditional measures: New insights from neurophysiological methods and market response modeling. *Journal of Marketing Research, 52*(4), 436–452.

H. Steven Scholte He holds a degree in the field of biology and psychology and a doctorate in medicine. Since 2005, he is a faculty member and, currently, an associate professor at the department of brain and cognition of the University of Amsterdam. His research focuses on the neural basis of visual information processing. He is an expert in modeling these processes with state-of-the-art machine learning techniques, and he runs a research group focused on making models of behavior that are inspired on neuroscience and constrained by neuroimaging studies. Apart from his academic career, he is the cofounder, and since 2019 the general director, of Neurensics BV.

Andries van der Leij He holds a degree in psychology. His graduate studies focused on the relationship between individual differences in the brain and behavior. He is an expert in the field of neuro-economics and social neuroscience, and he teaches in these areas at the University of Amsterdam. Apart from his work at the university, he is currently a partner at Neurensics BV where he is responsible for product development and innovation.

Victor A. F. Lamme He holds a degree and doctorate in medicine. After obtaining his PhD, he was a postdoc at MIT. He is currently full professor of psychology at the University of Amsterdam, and his research focuses on conscious and unconscious processing and consciousness research. He is a cofounder of Neurensics BV where he is the chief science officer.