Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2017 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

Summer 7-19-2017

Predicting Popularity of Hedonic Digital Content via Artificial Intelligence Imagery Analysis of Thumbnails

Stefan Cremer
University of Cologne, stefan.cremer@uni-koeln.de

Predicting Popularity of Hedonic Digital Content via Artificial Intelligence Imagery Analysis of Thumbnails

Completed Research Paper

Stefan Cremer
University of Cologne
Pohligstr. 1, 50969 Cologne, Germany
stefan.cremer@uni-koeln.de

Abstract

Hedonic digital content backs a wide variety of business models. Yet, due to its experience good nature, consumers cannot assess its value before consumption. To overcome this obstacle, thumbnail images are frequently employed to provide an experience of content, and trigger views and sales. In spite of fragmented evidence from human-computer interaction research, thumbnails largely constitute a black box for research and practice. This research aims to fill this gap and asks: How and why do basic, conceptual and social features of thumbnail images affect popularity of hedonic digital content? To answer the question, we employ artificial intelligence imagery analysis to test and confirm a variance model against evidence from 400,000 YouTube videos. Our findings entail important theoretical contributions to visual perception in online contexts. In addition, this research proposes artificial intelligence imagery analysis as a new and fruitful research method for the largely visual information systems discipline.

Keywords: Artificial intelligence, visual perception, human-computer interaction, hedonic digital content

Introduction

Innovative technologies have triggered today's omnipresence of hedonic digital content, such as videos, music or e-books. Digital content is commonly considered an experience good that has to be consumed before its intrinsic or monetary value can be assessed (Nelson 1970). Still, companies have to provide consumers with information cues about digital content, to market it via advertisement- or direct-payment-based business models. Thumbnail images (such as covers of movies, videos, music albums, or e-books) serve as the dominant information cue for digital content, as they maximize the amount and richness of information that can be communicated while minimizing the time needed for assimilation to as low as milliseconds, as opposed to text or video.

Thumbnail images become even more important in the advent of smartphones and tablets, since their touch-based control entails affordances for simplified and more efficient interfaces. Because human working memory is limited (Kahneman 1973), thumbnail images lower information overload compared to text-heavy interfaces, as they target the subconsciously working part of the visual sense with its high bandwidth for information processing. The continuously growing generation of digital natives tends to favor image-based interfaces over text-based interfaces (Djamasbi et al. 2016), which is for instance observed in social media where the image-based Instagram outperforms popularity of the multimedia-based Facebook for the young generation (Chowdhry 2014).

A problem with images in general and thumbnails for digital content in particular is that they largely constitute a black box, and still drive important and tangible outcomes such as systems' adoption and use and digital content consumption and sales. From a practice perspective, thumbnail images are commonly used on a best practice basis for driving consumption and sales – without exactly knowing if, how and why they work.

Theory nourished from information systems (IS) and media psychology research provides at most fragmented insights on the link between (thumbnail) images and outcomes such as the consumption of digital content. At the same time, IS researchers only recently called for an increased use of images as a data source, since the field is "overwhelmingly visual in nature" (Andrade et al. 2015).

To date, it was barely possible to dig deeper into the contentual and semantic level of large amounts of pictorial data and systematically analyze them with regard to the outcomes they trigger. Recent advancements in artificial intelligence made it possible for the first time to analyze large amounts of images with regard their features – the basic ones (dominant colors, shapes, or symmetry), textual, conceptual (e.g. important topics represented) and human features (e.g. faces and emotion), with a precision that recently exceeded the rating and classification precision of human raters (He et al. 2015). In addition, artificial intelligence imagery analysis seems to be a promising new method for IS research, as it – compared to research with human raters – overcomes typical human-related biases, allows for comprehensive pre-testing, simplifies longitudinal research designs that require collecting data over weeks or months, and reduces cost and time for data collection by a factor of 100 to 1000.

This research aims to make a first step towards closing the identified theoretical and methodological gaps, and demonstrating the applicability of artificial intelligence imagery analysis. Its principal research question is: How and why do basic, conceptual and social features of thumbnail images affect popularity of hedonic digital content?

To answer the question, we first conduct an extensive literature review and deduce a two-dimensional space for theory development on imagery features, spanned by the orthogonal dimensions 'classes of visual information cues' and 'attributes of visual information cues'. Based thereupon, we develop a variance model to predict consumption of hedonic digital content from visual information cues in thumbnail images. Using Microsoft Cognitive Services, a set of artificial intelligence API functions, we test the model against evidence from a random sample of 400,000 YouTube video thumbnails.

With this research, we aim to make a twofold contribution. First, with regard to IS theory, we offer insights on the link between the contentual and semantic level of imagery data and consumption or use as some of the most studied outcomes in IS research. Second, with regard to IS methodology, we seek to demonstrate the applicability and fruitfulness of artificial intelligence imagery analysis for IS research.

Human visual information processing: foundations, outcomes and measurement

Foundations of Human Visual Information Processing

Research on human visual information processing is an interdisciplinary endeavor, informed by medical sciences, neuroscience, biology, psychology, computer science and information systems.

Human visual information processing takes place in a complex system formed by the eyes, the optic nerve and the visual cortex (Hubel 1988). Through this system, humans perceive and evaluate their surroundings, commonly conceptualized as a scene with objects that possess characteristics (Henderson and Hollingworth 1999). Dominant research streams can be distinguished according to their bottom-up or top-down perspective on visual information processing (Eckhorn 2002). Interpretive, bottom-up theories regard a scene as triggering attention, cognition, emotion and action (Scott 1994), while the constructivist, top-down theories take the reverse perspective.

From a bottom-up perspective, visual information processing starts with the reflection of light from objects in a scene that is detected by the eyes' photoreceptors. They transform the light stimulus into an electric signal, emitted to the visual cortex. The visual cortex is modular in nature, so signals traverse different modules in the process of perception and cognition. Low-level processing comprises the detection of edges and shapes. High-level processing compares the detected shapes or combinations thereof with representations stored in episodic or declarative memory, and eventually triggers cognition, emotion and action. During their evolution, humans developed highly specialized modules for certain life- and survival-related tasks, such as processing human faces and facial cues for emotions (Snowden et al. 2012).

Visual information processing is contingent on working memory and attention. The limitation of working memory (Miller 1956) becomes particularly apparent in what dual-process theories (e.g., Kahneman 2011) call the conscious, effortful and slow system 2 of information processing. It is contrasted to the subconscious, automatic and fast system 1. Both systems can process the same visual information in parallel, but come to different conclusions (Kahneman 2011), e.g. when a dialog partner has a false smile, which is interpreted as “friendly and straightforward” by the conscious system 2 and as “hiding something important” by the subconscious system 1. The inherent, even though unequal working memory limitations in system 1 and 2 explain why human visual information processing is heuristic instead of exact and deterministic. Since working memory is limited, selective attention becomes another important contingency of visual information processing. In line with the top-down and bottom-up view, attention can either be triggered by an object or characteristic in the scene (e.g., a tiger appearing) or by a cognition or emotion (e.g., centering the attention on a compound in a zoo based on the expectation – a cognition – that a tiger is resident therein).

Outcomes and Measurement of Human Visual Information Processing: An Information Systems Research Perspective

Information systems research has investigated attitudinal and behavioral outcomes of processing information stored in images. Most of the work is descriptive in nature while a minority takes a prescriptive, design-science focus.

Attitudinal outcomes from visual information processing, following emotion or cognition, comprise the sensation of appeal, enjoyment, pleasure and preference (Cyr et al. 2009; Deng and Poole 2010; Hassanein and Head 2007; Steinmann et al. 2014), the sensation of social presence (Cyr et al. 2009), the sensation of helpfulness or relevance (Jiang and Benbasat 2007), and purchase intention (Steinmann et al. 2014). Behavioral outcomes comprise approach-avoidance behavior towards websites (Deng and Poole 2010) and bidding behavior in Internet auctions (Rafaeli and Noy 2005).

Methodologically, information systems research on human visual information processing is partly qualitative, i.e. analyzing interview data (Cyr et al. 2009), but mainly quantitative. The dominant quantitative approach is the experiment. Experimental approaches can be further differentiated into traditional, neuroimaging and psychophysiological approaches. In experiments following the traditional approach, participants typically view imagery stimuli on a screen and answer questions or make choice to conclude on attitudinal or behavioral outcomes (Jiang and Benbasat 2007; Rafaeli and Noy 2005; Steinmann et al. 2014). Neuroimaging studies use data obtained via electroencephalography (EEG; Gregor et al. 2014) or functional magnetic resonance imaging (fMRI; Benbasat et al. 2010; Riedl et al. 2011). In addition, functional near infrared spectroscopy (fNIR) has

been suggested as a third neuroimaging approach (Gefen et al. 2014). The dominant psychophysiological approach is eye tracking (Cyr et al. 2009; Djamasbi et al. 2010). Quantitative, experimental approaches are complemented by surveys (Cyr et al. 2009).

Thumbnail visual information cues impacting hedonic digital content popularity: a model

Conceptual Foundations and Boundaries

The model's unit of analysis is the individual viewing a set of thumbnails for different units of content to make a consumption decision. Thus, on a conceptual level, the model needs to specify its understanding of a thumbnail and the choice process.

With respect to the thumbnail, we stick to the dominant conceptualization in the literature and regard a thumbnail as a scene with objects that possess characteristics (Henderson and Hollingworth 1999). With respect to the choice process, we restrict our model to hedonic content (consumed for the sake of enjoyment and entertainment) and exclude functional content (consumed to achieve a certain goal). In contrast to functional content, choice of hedonic content is mainly triggered by emotion (as opposed to cognition). A combination of the circumplex model of emotion (Russel 1980) and reversal theory (Apter 1982) has been suggested as a theoretical lens for differentiating between functional, goal-oriented or 'telic' consumption and hedonic, non-goal-oriented or 'paratelic' consumption in digital contexts (Deng and Poole 2010). The circumplex model operationalizs emotion via two orthogonal dimensions: arousal (the 'activation' of an emotional state) and valence (the direction of the emotion; Russel 1980). According to reversal theory, in functional consumption, low levels of arousal are perceived as pleasant (assuming that a sensation of pleasantness serves as basis for a consumption decision), while hedonic consumption entails a para-telic (i.e., non-directed) state of mind associated with perceiving a high level of arousal as pleasant (Apter 2007; for an application to information systems, see Deng and Poole 2010). Thus, while both arousal and valence trigger emotions that trigger hedonic consumption, valence is of particular importance.

For a model on the impact of visual information cues on hedonic digital content popularity, we propose to locate dependent variables in a two-dimensional space spanned by two orthogonal dimensions: classes of visual information cues and attributes of visual information cues.

Dimension 1: Classes of Visual Information Cues

We employ a top-down, mono-hierarchic classification to build classes of visual information cues.

For the first classification level, we follow the distinction of human information processing into lower- and higher-level processing, where lower-level processing is more on the perception side and higher level processing is more on the cognition side. Lower-level processing comprises perceptions of contrast, color and shapes. On the first level of the classification, we thus differentiate between low-level cues (such as contrast, color, shapes) and high-level cues, covering all higher-level perceptions that are mentally constructed from lower level perceptions and episodic or declarative memory.

For high-level cues, relevant research – particularly in information systems – typically regards either conceptual cues (Jiang and Benbasat 2007; Speier 2006) or social cues (Cyr et al. 2009; Djamasbi et al. 2012; Hassanein and Head 2007). Furthermore, because of the involvement of different modules of the visual cortex, perception and cognition and the resulting attitude or behavior are contingent on whether a representation contains lifeless objects or living beings, particularly humans. We therefore propose to differentiate high-level cues into conceptual and social cues as sub-classes.

For the third and lowest level of the classification, we propose imagery cues and textual cues as sub-classes for conceptual cues. Imagery cues cover all lifeless things, such as a chair or a stone. Textual cues historically first resembled imagery cues till, over the millennia, script evolved as an own form of communication. Textual cues are considered as distinct from imagery cues, still are processed similarly in the visual cortex. In images, particularly in thumbnails, text is frequently used, e.g. to complement conceptual representations of imagery nature.

We further propose to differentiate social cues into social presence and emotional presence. Social presence refers to the presence of living beings, such as animals or humans, while the presence is emotionally neutral or emotions cannot be detected as a whole body is only visible from distance. Emotional presence refers to whether the living being facially communicates proper emotional states, such as anger, contempt, disgust, fear, happiness, sadness or surprise.

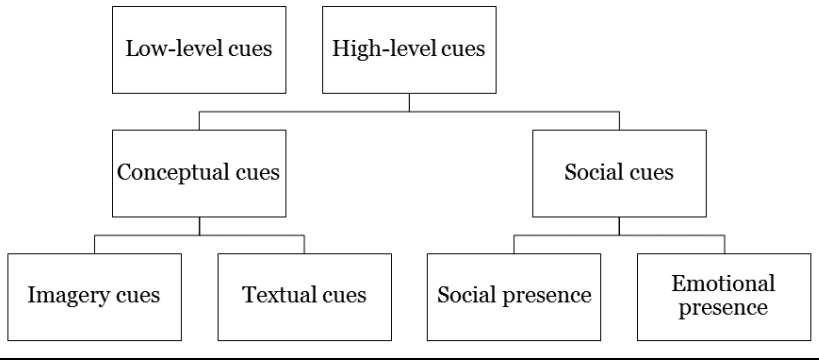

Figure 1. A Classification of Visual Information Cues

Dimension 2: Attributes of Visual Information Cues

For the second dimension, we propose to differentiate between complexity, valence, clarity, uniqueness of visual information cues for each of the classes in dimension 1.

Complexity

Creating a thumbnail for hedonic digital content requires balancing between simplicity and complexity. On the one hand, the spatial limitations of a thumbnail require strong condensation of the content into a simple pictorial summary. On the other hand, the more complex a thumbnail is, the more valid and vivid it communicates the content. Past information systems research provides conflicting rationales for both visual complexity and visual simplicity with regard to their conduciveness for consumption.

Some work provides evidence in favor of complexity, identifying it as positively impacting user behavior on websites, mediated by arousal (Deng and Poole 2010). This view is complemented by reversal theory which suggests that in para-telic, hedonic consumption situations, high levels of arousal are perceived as particularly pleasant, in contrast to functional consumption (Apter 1982).

A conflicting view finds that particularly younger generations have a lower tolerance for visual complexity (Djamasbi et al. 2011). This is in line with cognitive load theory (Sweller 1988), for which related research has found that users respond to complexity with preference for simplicity and becoming passive in text-centered online environments (Jones et al. 2004).

We thus propose the following pairs of opposing hypotheses for complexity:

$H_{1a(b)}$: Visual complexity in form of the number of high-level classes present positively (negatively) impacts consumption of hedonic digital content.

$H_{2a(b)}$: Imagery concept complexity positively (negatively) impacts consumption of hedonic digital content.

$H_{3a(b)}$: Textual complexity positively (negatively) impacts consumption of hedonic digital content.

$H_{4a(b)}$: Social complexity positively (negatively) impacts consumption of hedonic digital content.

$H_{5a(b)}$: Emotional complexity positively (negatively) impacts consumption of hedonic digital content.

Valence

Valence is the directional, positive or negative component of emotion (Russel 1980). There are two opposing views on the impact of valence on hedonic consumption.

One stream of literature proposes that negative content triggers more consumption than positive content. Baumeister et al. (2001) investigate negativity bias as the underling mechanism. In their view, preference for negativity was adaptive during human evolution. Humans who directed more attention to negative than to positive information in their environment had a higher probability to survive (Baumeister et al. 2001). In the contemporary world, the environment that shaped our mental preferences has changed, while the preferences have not. Complementary research to the "bad is good" view shows that of people who can choose between good and bad news, 77%-88% consume bad

news first (Marshall and Kidd 1981). On websites, people direct greater attention to and show better memory performance for negative stimuli (Kaspar et al. 2015).

A contrasting view on the impact of valence on hedonic consumption is provided by emotional contagion theory (Hatfield et al. 1993). According to emotional contagion theory, people unconsciously mirror emotions they observe in the environment. This effect is not limited to face-to-face interaction, but can also be present in technology-mediated interaction, and even if this interaction is asynchronous and/or textual (Kramer et al. 2014). According to neuroscience research, mirror neurons are responsible for the mirroring of emotions (van der Gaag et al. 2007). For online environments it was shown that positive content triggers impulse behavior on websites via perceived enjoyment as mediator (Parboteeah et al. 2009).

With regard to the relevant classes of visual information cues, we propose to include emotional valence into the model and exclude conceptual cues and social presence. For emotional valence communicated via facial images, there is robust evidence that they evoke emotions in a similar vein across individuals. We exclude conceptual cues and social presence from the valence dimension of the model, as their valence is highly contingent on the consumption situation and individual preferences.

We thus propose for valence:

$H_{6a(b)}$: Emotional valence positively (negatively) impacts consumption of a hedonic digital content.

Concreteness

While complexity is a function of the number of elements that are present for a certain category, concreteness refers to how concretely or vaguely an object is represented in a scene. We define concreteness as the sensory richness of a stimulus (Steuer 1992). With regard to concreteness, the literature offers two opposing directions for the effect on consumption of a hedonic digital content.

The stream arguing in favor of concreteness states that concreteness, or – in the words of Steinmann et al. (2014) – the quality of a presentation, is an antecedent for visual complexity (Deng and Poole 2010). Visual complexity triggers arousal and thereby positively impacts user behavior on websites (Deng and Poole 2010).

An opposing view takes the perspective on hedonic digital content as artwork. It is well known that art lives from and plays with vagueness and ambiguity (Mamassian 2008). A novel or movie, for example, would lose much of its attractiveness if the characteristics and intentions of all protagonists are concretely stated in the beginning. This view is further supported by communication theory (Reddy 1979; Sperber and Wilson 1995), according to which humans use two communication channels: explicit code and context-mediated information. The more information is explicitly and concretely encoded into a thumbnail, the less a viewer needs to rely on context-mediated information. On the other hand, the 'space' for explicit information in the thumbnail is very limited. Thus, and as for example done in advertising, a producer might only make vague visual hints on the content, that require the viewer to use a commonly shared context (e.g. a country's culture) to create a mental image that is complete. In this vein, the producer can communicate much more information and might ultimately make the content more attractive. We thus propose for concreteness:

$H_{7a(b)}$: Image resolution positively (negatively) impacts consumption of hedonic digital content.

$H_{8a(b)}$: Conceptual concreteness positively (negatively) impacts consumption of hedonic digital content.

Uniqueness

Uniqueness refers to whether a visual information cue in a thumbnail is rather unique or non-unique within a content category or on a platform.

One stream of research suggests that uniqueness of visual information cues in thumbnails is favorable for consumption. Novel information cues require more initial processing (Gleasure 2014) and thus might better capture the attention of viewers. Second, novelty is an antecedent of appraisal (Ellsworth 2013).

A contrary view is provided by the notion that taste for hedonic content continuously homogenizes and gravitates towards uniformity (Fu 2010). In line with this view, viewers might have learned through socialization to favor averageness toward uniqueness. Also, viewers can process non-unique and better known visual cues in images faster (Gleasure 2014), which is associated with lower

cognitive load (Sweller 1988), playing to younger generations who exhibit impatient viewing behavior (Djamasbi et al. 2016). We thus propose for uniqueness:

$H_{9a(b)}$: Color scheme uniqueness positively (negatively) impacts consumption of hedonic digital content.

$H_{10a(b)}$: Conceptual uniqueness positively (negatively) impacts consumption of hedonic digital content.

Control Variable

The age distribution of the users of a content platform might not reflect the age distribution of the overall population. Plus, viewers may exhibit a viewing behavior that favors content with protagonists from their age group. We thus control the average age of the protagonists pictured in a thumbnail:

$H_{11a(b)}$: The average age of the protagonists shown in a thumbnail positively (negatively) impacts consumption of hedonic digital content.

Methodology

Research Design, Method and Context

To answer the research question, we employ an observational, cross-sectional study design. As research method, we analyze the visual content of thumbnail images for a sample of 400,000 YouTube videos via artificial intelligence imagery analysis, and relate the findings to views of the respective videos. Artificial intelligence imagery analysis is an appropriate method to test our research model, since deep neural networks mimic the functioning of the human brain and visual perception and because deep neural networks are initially fed and calibrated with data from human raters.

To conduct artificial intelligence imagery analysis, we choose the Microsoft Cognitive Services APIs for computer vision, faces and emotion. These APIs detect, among others, imagery concepts, faces and emotion in faces within pictures. Their classification precision had already exceeded classification precision of human raters two years prior to our study (He et al. 2015). More technical descriptions for these particular services are provided by He et al. (2014) for the Computer Vision API, by Chen et al. (2013, 2014) for the Face API, and Yu and Zhang (2015) for the Emotion API.

Operationalization of Variables

We operationalize the independent variables as follows:

- *Visual complexity*: A number between 0-3 indicating how many high-level classes are present at all: conceptual cues (imagery), conceptual cues (textual), social cues (= presence of faces)
- *Imagery concept complexity:* A number indicating how many imagery conceptual cues are present. Examples from the dataset are 'sport', 'road', 'grass', and 'plane'.
- *Textual complexity*: A number indicating how many characters of text are present.
- *Social complexity*: A number indicating how many faces are present.
- *Emotional complexity*: A number indicating how many non-neutral emotions are present over all faces in the picture. Non-neutral emotions are anger, contempt, disgust, fear, happiness and sadness (as opposed to 'neutral' and 'surprise', which are also detected by the API). An emotion is counted if it has a strength > .05 (1 = strongest presence of the emotion).
- *Emotional valence*: We operationalize this variable via four sub-variables
    - *Average negative (positive) valence:* the valence of the strongest negative (positive) emotion for each face, cumulated for all faces and divided by the number of faces.
    - *Strongest negative (positive) valence:* the valence of the strongest negative (positive) emotion among all faces
- *Image resolution*: resolution of the thumbnail in megapixel
- *Conceptual concreteness:* Average confidence with which an imagery concept was detected (a confidence score between 0 and 1 is assigned to each concept)

- *Color scheme uniqueness:* One divided by the frequency a color scheme is present in the dataset. A color scheme is a pair [dominant background color, dominant foreground color].
- *Conceptual uniqueness:* One divided by the frequency an imagery concept is present in the dataset, cumulated and weighted for all concepts in a picture.
- *Average age:* The average age of the humans shown in the picture.

Data Collection

In a first step, we obtained a list of videos from the YouTube-8M dataset (YouTube 2016), an 8 million video random sample of YouTube videos that is available for researchers for the purpose of developing and improving systems for video analysis, e.g. for content tagging and classification. We then drew a random sub-sample of 400,000 videos. For these videos, we obtained metadata such as the number of views or the number of subscribers to the video's channel. In a second step, we sequentially submitted the 400,000 thumbnails to the Computer Vision API, Face API and Emotion API. For obtaining complementary data on text within the thumbnails, we processed all pictures with Tesseract, an open-source character recognition engine.

Results

Descriptive Statistics

Overall, the dataset comprises records for 395,764 thumbnails and videos from 2005-2015 (a small number of videos had been deleted since the release of the YouTube-8M dataset). The average video has 56,788 views and a rating of 4.37/5. In the thumbnails, there are overall 159 different color schemes. In 76% of the thumbnails, imagery analysis detected on average 3.22 imagery conceptual cues (such as 'sport', 'road', 'grass', and 'plane'; average confidence: 0.92) from a total of 1439 different concepts within the dataset. 8% of the thumbnails contain text, with an average length of 76 characters. 21% show at least one face. Among those, there are 1.52 faces on average, picturing persons with an average age of 31 years.

Multiple Linear Regression Analysis

Multiple linear regression was calculated to predict consumption of hedonic digital content from the independent variables. Table 1 lists the regression coefficients.

Model Term	Coefficient		t-value	Sig.
Intercept	69,871	***	10.067	.00
Visual complexity	-12,230	***	-0.913	.01
Imagery concept complexity	-18,439	n.s.	-1.533	.13
Textual complexity	-1,065	*	-1.678	.09
Social complexity	45,263	***	2.592	.01
Emotional complexity	33,217	*	1.633	.10
Emotional valence (avg. negative)	-4,668,783	*	-1.833	.07
Emotional valence (strongest negative)	3,807,328	*	1.727	.08
Emotional valence (avg. positive)	112,133	n.s.	0.349	.73
Emotional valence (strongest positive)	-72,024	n.s.	-0.238	.81
Image resolution	-27,231	***	-4.622	.00
Conceptual concreteness	-11,589	*	-1.636	.10
Color scheme uniqueness	27,360,573	***	3.846	.00
Conceptual uniqueness	274,795	n.s.	0.079	.94
Average age	-1,183	**	-1.960	.05

*** / ** / *: significant at level p < .01 / p < .05 / p < .1; n.s.: not significant

Table 1. Regression Coefficients

Conclusion and discussion

As expected, we find significant effects within all dimensions of our model. In the complexity dimension, a one-unit increase in visual complexity (to up to 3, if three high-level classes are present) is associated with a decrease in views by 12,000. Picturing one more imagery concept is associated with a decrease in views by 18,000, even though the effect just missed the significance threshold ($p = .13$). Adding one more character of text to the thumbnail is associated with a decrease in views by 1000. In contrast, picturing one more face in the thumbnail leads to an average increase in views by 45,000. An increase in emotional complexity, i.e. a one-unit increase in the number of non-neutral emotions shown across all faces, is associated with an increase in views by 33,000. Overall, evidence for the complexity dimension suggests that adding conceptual cues to thumbnails (imagery cues and textual cues) is, on average, detrimental for views, while adding faces and 'loading them with emotion' is positive for views. This supports the perspective of cognitive load theory (Sweller 1988) combined with the view that tolerance for visual complexity decreases in a world of information overload (Djamasbi et al. 2011; Jones et al. 2004). Imagery cues and particularly textual cues are c.p. harder to process than social cues, so they are associated with higher cognitive load. Our results suggest that when being confronted with a choice set of thumbnails, consumers just focus on the easy and fast to process social cues and rather disregard imagery and textual cues, and that this initial attention to thumbnails with social cues translates into viewing behavior. This view also disconfirms information systems theory stating that on websites, complexity leads to arousal and thereby triggers user behavior (Deng and Poole 2010). Our findings suggest, that just social complexity leads to arousal and drives user behavior – while complexity of non-social cues leads to passivity.

In the valence dimension, both effects for negative emotions are significant, while both effects for positive emotions are insignificant. If several faces are pictured, an increase in the average of their strongest negative emotion is associated with an increase in views. An increase in the strongest negative emotion among all faces, however, is associated with a decrease in views. Overall, both the conduciveness of 'average negativity' in emotion and the insignificance of the effects for positive emotions support the 'bad is good' hypothesis (Baumeister et al. 2001), in that negative emotions are better able to capture attention and thereby trigger behavior. The reversed direction between the two effects for negative valence suggest that particularly groups of people and the homogeneousness with regard to the negative emotion are associated with more views.

In the concreteness dimension, we find that an increase of the thumbnail resolution by one megapixel is associated with a decrease in views by 27,000. An increase of conceptual concreteness by 1, the maximum possible increase, is associated with a decrease in views by 12,000. Both findings support the 'art is vague and ambiguous' hypothesis (Mamassian 2008) and are in contrast to the assumption that quality triggers user behavior via the links of visual complexity and arousal (Deng and Poole 2010).

In the uniqueness dimension, color scheme uniqueness is associated with an increase in views, while the effect for conceptual uniqueness is insignificant. The direction of the first effect favors the hypothesis that novel information cues might impact attention and behavior, as they require more initial processing (Gleasure 2014).

Theoretical, Methodological and Managerial Contribution

Theoretical Contribution

Our findings make theoretical contributions in three areas. First, we extend and partly disconfirm a long accepted view in information systems research, namely that visual complexity on websites leads to arousal and thereby triggers user behavior (Deng and Poole 2010). Our findings suggest that it is important to differentiate between 'conceptual complexity' and 'social complexity' and that – at least in the context of our study, the latter is conducive for triggering user behavior, while the former is not. We also find partial support for the 'bad is good' hypothesis, confirming common wisdom that in hedonic consumption contexts from movies to news, visual information cues with positive valence rather do not play a role, while their negative counterparts capture attention and evoke user behavior. We also find some support for the 'vague and ambiguous is good' hypothesis, which is in line with the notion that 'quality' in hedonic consumption contexts does not mean everything has to be precisely and straightforward described, but that some ambiguity and open ends, particularly on the visual level, better capture attention and trigger user behavior.

Methodological Contribution

To the best of our knowledge, we are among the first to apply artificial intelligence imagery analysis in information systems research. With this paper, we seek to demonstrate the applicability and scope of the method and suggest its consideration for the methodological basket of IS research. We envision the method's applicability for a wide variety of IS research contexts, such as design science in human-computer interaction, the study of adoption and usage of information systems and technology, particularly with regard to the visual components, and the study of the visual-content-driven social media. In terms of rating precision (He et al. 2015), the reduction of common biases, the enabling of comprehensive pre-testing, the enabling of new longitudinal research designs, and the reduction of cost and time for data collection, artificial intelligence imagery analysis is a promising alternative or supplement to traditional research methods.

Managerial Contribution

With regard to practice, our findings suggest that mangers responsible for advertisement- or direct-payment-based business models around digital content should even more differentiate between the 'packaging' and the 'content'. If no one opens the package, even the best content is of no value, and this challenge aggravates in a world of information ubiquitousness and scarce attention.

More particular, managers should focus on the social component of the 'packaging'. Our results demonstrate that social cues are especially suitable to trigger attention and viewing behavior. If it is suitable for the content, showing a group of people is better than picturing a single face. Furthermore, and even if counterintuitive to some common wisdom, picturing a homogenous amount of negative emotion within a group might also be conducive for getting attention – in the sense of 'bad is good'. If the thumbnail shows imagery concepts, they should not be too concrete, but a bit vague instead, to grab the viewers' attention.

With regard to business process management, managers can further implement artificial intelligence imagery analysis into their processes or for their end-user platforms. For internal processes, the method can lead to recommendations for the design of content 'packaging' and give real-time feedback to design drafts. The same can be implemented into platforms where producers upload their own content, to build toolboxes and checklists for stronger embedding of the content into the platform via its 'packaging'.

Limitations and Further Research

As this research is exploratory in nature, future research is needed to dig deeper into the mechanisms that trigger hedonic content consumption within the boundaries of 'classes of visual information cues' and 'attributes of visual information cues'. Few results, e.g. the ambiguous results for negative emotional valence, should be further tested via an updated research design to increase the robustness of the model.

To demonstrate reliability of our model, future research should test it in other contexts, e.g. for book and DVD covers. Also, future research could differentiate between categories of content or investigate functional content to strengthen the boundaries of the model.

Finally, for some of the effects shown, causality might not only flow into one direction. It is conceivable that more successful producers of content create, on average, more sophisticated thumbnails. Future research could seek to conceptually separate to what degree semi-professional or professional producers of content use more visual features in their thumbnails that – according to our model – trigger consumption behavior.

References

Apter, M. 1982. *The Experience of Motivation: The Theory of Psychological Reversals*, London: Academic Press.

Apter, M. 2007. *Reversal Theory: The Dynamics of Motivation, Emotion and Personality*, Oxford: Oneworld Publications.

Baumeister, R., Bratslavsky, E., Finkenauer, C., and Vohs, K. 2001. "Bad is stronger than good," *Review of General Psychology* (5:4), pp. 323-370.

Benbasat, I., Dimoka, A., Pavlou, P., and Qiu, L. 2010. "Incorporating Social Presence in the Design of the Anthropomorphic Interface of Recommendation Agents: Insights from an fMRI Study," in *Proceedings of the International Conference on Information Systems (ICIS)*, Saint Louis, USA, December 15-18.

Chen, D., Ren, S., Wei, Y., Cao, X., and Sun, J. 2014. "Joint Cascade Face Detection and Alignment," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Zürich, Switzerland, September 6-12.

Chen, D., Cao, X., Wen, F., and Sun, J. 2013. "Blessing of Dimensionality: High Dimensional Feature and Its Efficient Compression for Face Verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, USA, June 23-28.

Cyr, D., Head, M., Larios, H., and Pan, B. 2009. "Exploring human images in website design: a multi-method approach," *MIS Quarterly* (33:3), pp. 539-566.

Cyr, D., Head, M., and Larios, H. 2010. "Colour appeal in website design within and across cultures: A multi-method evaluation," *International Journal of Human-Computer Studies* (68:1), pp. 1-21.

Deng, L., and Poole, M. 2010. "Affect in web interfaces: a study of the impacts of web page visual complexity and order," *MIS Quarterly* (34:4), pp. 711-730.

Andrade, A., Urquhart, C., and Arthanari, T. 2015. "Seeing for Understanding: Unlocking the Potential of Visual Research in Information Systems," *Journal of the Association for Information Systems* (16:8), pp. 646-673.

Chowdhry, A. 2014. "Survey Says Teenagers Prefer Instagram Over Facebook," *Forbes*, http://www.forbes.com/sites/amitchowdhry/2014/10/13/survey-says-teenagers-prefer-instagram-over-facebook/.

Djamasbi, S., Siegel, M., and Tullis, T. 2010. "Generation Y, web design, and eye tracking," *International Journal of Human-Computer Studies* (68:5), pp. 307-323.

Djamasbi, S., Siegel, M., Skorinko, J., and Tullis, T. 2011. "Online viewing and aesthetic preferences of generation y and the baby boom generation: Testing user web site experience through eye tracking," *International Journal of Electronic Commerce* (15:4), pp. 121-158.

Djamasbi, S., Siegel, M., and Tullis, T. 2012. "Faces and viewing behavior: An exploratory investigation," *AIS Transactions on Human-Computer Interaction* (4:3), pp. 190-211.

Djamasbi, S., Strong, D., Wilson, E. and Ruiz, C. 2016. "Designing and Testing User-Centric Systems with both User Experience and Design Science Research Principles," in *Proceedings of the Special Interest Group on Human-Computer Interaction*, San Diego, USA, August 11-14.

Eckhorn, R. 2002. "Neural principles of preattentive scene segmentation: Hints from cortical recordings, related models, and perception," in *Models of Neural Networks IV*, J. van Hemmen, J. Cowan, and E. Domany (eds.), New York: Springer, pp. 183-216.

Ellsworth, P. 2013. "Appraisal theory: Old and new questions," *Emotion Review* (5:2), pp. 125-131.

Fu, W. 2013. "National Audience Tastes in Hollywood Film Genres: Cultural Distance and Linguistic Affinity," *Communication Research* (40:6), pp. 789-817.

Gefen, D., Ayaz, H., and Onaral, B. 2014. "Applying functional near infrared (fNIR) spectroscopy to enhance MIS research," *AIS Transactions on Human-Computer Interaction* (6:3), pp. 55-73.

Gleasure, R. 2014. "Using distractor images in web design to increase content familiarity: a NeuroIS perspective," in *Proceedings of the International Conference on Information Systems (ICIS)*, Auckland, New Zealand, December 14-17.

Gregor, S., Lin, A., Gedeon, T., Riaz, A., and Zhu, D. 2014. "Neuroscience and a nomological network for the understanding and assessment of emotions in information systems research," *Journal of Management Information Systems* (30:4), pp. 13-48.

Hassanein, K., and Head, M. 2007. "Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping," *International Journal of Human-Computer Studies* (65:8), pp. 689-708.

Hatfield, E., Cacioppo, J., and Rapson, R. 1993. "Emotional contagion," *Current Directions in Psychological Science* (2:3), pp. 96-100.

He, K., Zhang, X., Ren, S., and Sun, J. 2015. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *Microsoft Research*.

He, K., Zhang, X., Ren, S., and Sun, J. 2014. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," in *Proceedings of European Conference on Computer Vision (ECCV)*, Zürich, Switzerland, September 6-12.

Henderson, J., and Hollingworth, A. 1999. "High-level scene perception," *Annual Review of Psychology* (50:1), pp. 243-271.

Jiang, Z., and Benbasat, I. 2007. "The effects of presentation formats and task complexity on online consumers' product understanding," *MIS Quarterly* (31:3), pp. 475-500.

Jones, Q., Ravid, G., and Rafaeli, S. 2004. "Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration," *Information Systems Research* (15:2), pp. 194-210.

Kahneman, D. 2011. *Thinking, fast and slow,* New York: Farrar, Straus and Giroux.

Kahneman, D. 1973. *Attention and effort*, Englewood Cliffs: Prentice-Hall.

Kaspar, K., Gameiro, R., and König, P. 2015. "Feeling good, searching the bad: Positive priming increases attention and memory for negative stimuli on webpages," *Computers in Human Behavior* (53), pp. 332-343.

Kramer, A., Guillory, J., and Hancock, J. 2014. "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences* (11:(24), pp. 8788-8790.

Mamassian, P. 2008. "Ambiguities and conventions in the perception of visual art," *Vision Research* (48:20), pp. 2143-2153.

Marshall, L., and Kidd, R. 1981. "Good news or bad news first?," *Social Behavior and Personality* (9:2), pp. 223-226.

Miller, G. 1956. "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological Review,* (101:2), pp. 343-352.

Nelson, P. 1970. "Information and consumer behavior," *Journal of Political Economy* (78:2), pp. 311-329.

Parboteeah, D., Valacich, J., and Wells, J. 2009. "The influence of website characteristics on a consumer's urge to buy impulsively," *Information Systems Research* (20:1), pp. 60-78.

Rafaeli, S., and Noy, A. 2005. "Social presence: influence on bidders in internet auctions," *Electronic Markets* (15:2), pp. 158-175.

Reddy, M. 1979. "The conduit metaphor – a case of frame conflict in our language about language," in *Metaphor and Thought*, A. Ortony (ed.), Cambridge: Cambridge University Press, pp. 284–324.

Riedl, R., Mohr, P., Kenning, P., Davis, F., and Heekeren, H. 2011. "Trusting humans and avatars: Behavioral and neural evidence," in *Proceedings of the International Conference on Information Systems (ICIS)*, Shanghai, China, December 4-7.

Russel, J. (1980). "A circumplex model of affect," *Journal of Personality and Social Psychology* (39:6), pp. 1161-1178.

Scott, L. 1994. "Images in Advertising: The Need for a Theory of Visual Rhetoric," *Journal of Consumer Research* (21:2), pp. 252-273.

Snowden, R., Thompson, P., and Troscianko, T. 2012. *Basic Vision: An Introduction to Visual Perception*, Oxford: Oxford University Press.

Speier, C. 2006. "The influence of information presentation formats on complex task decision-making performance," *International Journal of Human-Computer Studies* (64:11), pp. 1115-1131.

Sperber, D., and Wilson, D. 1995. *Relevance: Communication and Cognition*, Oxford/Cambridge: Blackwell Publishers.

Steinmann, S., Kilian, T., and Brylla, D. 2014. "Experiencing Products Virtually: The Role of Vividness and Interactivity in Influencing Mental Imagery and User Reactions," in *Proceedings of the International Conference on Information Systems (ICIS)*, Auckland, New Zealand, December 14-17.

Steuer, J. 1992. "Defining virtual reality: Dimensions determining telepresence," *Journal of Communication* (42:4), pp. 73-93.

Sweller, J. 1988. "Cognitive load during problem solving: Effects on learning," *Cognitive Science* (12:2), pp. 257-285.

van der Gaag, C., Minderaa, R., and Keysers, C. 2007. "Facial expressions: what the mirror neuron system can and cannot tell us," *Social Neuroscience* (2:3-4), pp. 179-222.

Yu, Z., and Zhang, C. 2015. "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, Seattle, USA, November 9-13.