Following the viewers: Investigating television drama engagement through skin conductance measurements

Andreas Gregersen[a,*], Birger Langkjær[a], Lene Heiselberg[b], Jacob Lyng Wieland[b,c]

[a] Department of Media, Cognition and Communication, University of Copenhagen, Karen Blixens Plads 8, 2300 Copenhagen S, Denmark
[b] DR Research, Emil Holms Kanal 20, 0999 Copenhagen C, Denmark
[c] BBC Marketing and Audiences, BBC Broadcast Centre, Media Village, 201 Wood Lane, W12 7TP, London, United Kingdom

ARTICLE INFO

ABSTRACT

This article describes an empirical study of viewer emotional engagement with the first episode of the television fiction series *Follow the Money* (DR, 2015), a crime melodrama. The study combines applied and academic audience research and investigates real-time viewer responses by measuring and analysing skin conductance (SC) responses as an indicator of emotional arousal. The article outlines a framework for analysing emotion and arousal in television viewer engagement. It then describes SC measurements and the difference between tonic and phasic measurements. Two primary SC measures that indicate emotional engagement are derived, i.e. average phasic skin conductance response and the average number of emotionally arousing events pr. second. In our findings, both of these measures indicate a slight linear upwards trend and qualitative data is brought in to contextualize the quantitative findings. The study demonstrates how SC measures, formal analysis, and qualitative methods can be combined and the article ends with an outline of how future studies might adopt parts of the research design to further integrate applied audience research, audience research in media psychology, and cognitive film theory.

1. Introduction

It is a common assumption that viewers are drawn to fictional entertainment because of the emotions evoked. Most studies of this phenomenon, however, are either largely theoretical or based primarily on viewer self-reports. This study outlines an empirical approach that operationalizes emotional gratification as real-time emotional arousal as part of an ongoing effort to integrate three related perspectives relevant for the overall domain of audience research: audience research in media psychology, cognitive film theory, and applied audience research. The study was a pilot collaboration between two audience researchers working at Danish Broadcasting Corporation ('Danmarks Radio', hereafter DR) and two researchers working predominantly within the area of cognitive film theory.

We begin by outlining the common research area of audience emotional gratifications with a brief review of key points from media psychology in order to synthesise a theoretical framework. We tie this framework to that of cognitive film theory with a focus on character engagement and emotions. We then outline the organizational context for audience research in DR and end the section by describing our operationalization, our research questions, and our hypotheses. The next two sections describes the methodology and main quantitative findings, which are followed by a discussion which brings in 1) findings from the qualitative component of the study and 2) commentary on the content features which are possibly central to eliciting and maintaining viewer emotions. We end by

---

discussing some outstanding issues related to our chosen measures and segmentation procedures and outline areas for further research. We see our contribution as both theoretical and methodological: Most of our findings are tentative, but the study as a whole has implications for further work on media and emotional engagement in both applied and academic settings.

2. Theoretical overview

2.1. Entertainment and emotional gratifications: a tri-partite framework

Why do people watch television and what happens when they watch it? Many academic perspectives investigate these broad questions, but it has long been the consensus within psychology of entertainment that a main factor is the emotional payoff (see e.g. Bartsch, 2012; Tan, 1996, 2008; Vorderer & Hartmann, 2009), as summed up by Bartsch (2012):

"Emotions are often assumed to be the heart of media entertainment, be it in the form of movies, novels, television programs, music videos, or computer games. Entertainment audiences want to have a good laugh, they want to be kept at the edge of their seats, or be moved to tears" (p. 267–68).

We will assume that this is true, both across media and in the case of television: People watch television entertainment to experience emotions, and successful television entertainment delivers these experiences.

To start, we will outline some fundamental assumptions from media psychology regarding emotions and entertainment (for reviews see Bartsch & Viehoff, 2010; Bartsch, 2012; Tan, 2008; Vorderer & Hartmann, 2009). Emotional engagement in television fiction plausibly involves at least three distinct levels of processes. The first is direct and immediate emotional episodes, experienced as pleasant or unpleasant emotions, such as laughter or excitement etc. The second level involves short-term re-appraisal by way of meta-cognitive and meta-emotional processes, which regulate emotion (Frijda, 1986; Gross, 1999; Lazarus, 1991): viewers may thus react positively or negatively based on their immediate preferences for, say, laughing or being thrilled or scared.[1] In addition to this, a third level involves a further re-appraisal of media-related experiences as particularly meaningful over the longer term, later in or after the viewing process. This third level is a more complex and longer-term process termed *appreciation* by Oliver and Bartsch (2010). The present study builds on this tri-partite framework of 1) primary responses, 2) short-term re-appraisal and 3) longer term, complex re-appraisal, but the focus is on the first level, i.e. primary responses as immediate and intrinsically gratifying emotional aspects of television viewing. More operationally, the study assumes that observed increases in arousal during television viewing are indicators of the aforementioned primary emotional responses, which form the basis for the latter two levels.

2.2. Arousal as an indicator of emotional gratifications

Arousal is generally seen as a key dimension of emotional responses. Although considerable disagreement exist within the study of emotion, one of the dominant models of emotions within media psychology (Ravaja, 2004) and arguably within psychology of emotions in general (see Russell & Barrett, 1999; Russell, 2003, 2009) is the two-dimensional model, which distinguishes between valence and arousal as the key dimensions of emotional experience. The present study is aligned with this dimensional model.

It may be readily observed that the common-sense meaning of entertainment implies positive valence, but it has long been argued that emotions with negative valence, e.g. sadness, may be experienced as pleasing through the short-term and longer term re-appraisal processes described in the model above. Several theories on mood management and sensation seeking also hold that mild arousal in itself, regardless of valence, is gratifying. Recent studies have investigated the relationship between arousal and emotional gratifications further and Bartsch (2012) has uncovered seven stable emotional gratifications, i.e. fun, thrill, empathic sadness, contemplation, character engagement, social sharing and vicarious release. All of these components except fun were found to correlate with increases in arousal – but it should be noted that Bartsch measured arousal through self-report and as a single aggregated score. Given the tri-partite model and these findings, one would expect arousal to be a plausible indicator of many primary emotional responses to audio-visual fiction. One might also hypothesize that retrospective arousal is connected to arousal states during viewing, in that later self-reports plausibly piggyback on similar, if not identical, arousing emotions experienced during the actual viewing. Thus, even though immediate arousal is not a precise indicator of specific emotional gratifications it is nevertheless a plausible indicator of both experienced emotions in general as well as emotional gratifications specific to engagement with fiction.

2.3. Character engagement as the main link from screen to viewer

Having outlined this framework for emotional gratifications, we will now turn to the issue of how and why these emotions arise. In this section, we will outline a theoretical approach to this issue based in cognitive film theory. This approach shares central concerns with media psychology but offers a particular take on fictional engagement which one might refer to as *the primacy of character, conflict, and emotion*: emotions are deemed central to fiction, and a key driver of such emotions are characters and conflict.[2]

---

[1] Bartsch (2012) distinguishes between what she calls direct and indirect gratifications, which accords well with Vorderer and Hartmann's (2009) distinction between primary appraisal responses and a secondary re-appraisal of primary responses.

[2] This perspective is not unique to cognitive film theory; this broadly Aristotelian approach to analysis of fiction is also dominant in modern day screenwriting literature (see Thompson (2003)).

Key principles of cognitive film studies can be summarized as follows. First, viewers seek out audio-visual cues that allow them to build a reasonably coherent story (Bordwell, 1986, 2006) and this cognitive construction work is enabled by narrative design. Emotional engagement is a fundamental component of the viewing experience (see Carroll, 2003; Grodal, 1997; Plantinga, 2009; Tan, 1996), and characters are a focal point for cognitive and emotional engagement: In line with the Aristotelian tradition, characters are the main gateways for viewers to enter the world of fiction and become engaged by the story.[3] Typically, a small number of characters function as the main characters, and the narration is aligned with these characters. The result is a *structure of sympathy and antipathy* towards fictional characters evoked by the filmic narration (Plantinga, 2009; Smith, 1995): We care about and for the fictional characters and their goals, dreams, fortunes and mishaps. In short, this theoretical framework assumes that a designed focus on the characters and the resulting emotional engagement is fundamental to engagement with audio-visual fiction in general. We will now outline how similar issues have been researched empirically within applied audience research in the broadcast television sector.

3. Organizational context: DR audience research

Over the last ten years, the Danish Broadcasting Corporation (DR) has produced several successful television fiction series, such as *The Killing*, *Borgen*, and *The Legacy*. Like many other production organizations within this field, DR is acutely aware of audience responses in relation to their television series and, as a result, DR has established a laboratory for testing new methods for applied audience research. The laboratory was established in 2012, and over its lifetime, it has employed a mixture of qualitative and quantitative methods. Early audience test showings conducted by the lab have become a stable component of the in-house production of television series, especially the highly successful flagship shows mentioned above.

A typical viewer evaluation study at DR audience research is conducted in the following manner: The first two episodes of a television series in production (the "third edit"-version) are delivered to the laboratory, where the episodes are shown to a relevant sample of the intended audience and viewer responses are elicited and analysed. Summaries are produced to give relevant and actionable feedback to the Head of Fiction and the Channel Controller. Over the years, the researchers in the lab have developed a research battery tailored specifically to testing of responses to fiction series. Three specific aims of this research battery are the most relevant here: First, to identify the appeal (or lack of same) of the main characters and the possibility for viewer identification with these characters. Second, to identify possible key scenes in terms of dramatic appeal – or lack of same – in a bottom-up manner. Third, to probe the impact and appeal of previously identified key scenes. To meet these aims, the test battery includes both focus groups (with audio-visual stimulus material) and individual semi-structured interviews as well as several quantitative instruments, typically small open-ended questionnaires.[4]

In recent years, the focus in the lab has been on incorporating real-time audience feedback measurement methods which do not depend upon self-reports, i.e. methods such as eye tracking, measurements of brain activity through electroencephalography (EEG), and measurements of psychophysiological changes via electro-dermal activity (EDA). Such real-time feedback methods are especially important for television audience research, since the marketplace of viewership is competitive in real-time: If viewers are bored, they may switch immediately to another product. The possibility to identify "dead periods" has thus been a key draw for adoption of real-time measurement methods. In addition to these overall concerns, it should be noted that due to the embedding within the actual production process timeline, a fast turnaround time is crucial in an organization like DR. A typical evaluation of two television series episodes will thus be conducted within one week only, and this single week covers data collection, data consolidation, data analysis, interpretation and summarizing as well as final feedback delivered. As a direct consequence, time is truly of the essence at all stages in the research process.

Given this set of constraints, EDA measurements in the form of skin conductance (SC) measurement were chosen for the present study. EDA measurement is fairly common in media research (Kivikangas et al., 2011; Ravaja, 2004) and is increasingly used in applied settings (for examples, see Fleureau, Guillotel, and Orlac, 2013; Silveira, Eriksson, Sheth, and Sheppard, 2013; Wang et al., 2016). DR's reasons for adopting EDA technology were based primarily on considerations of cost, ease of use, and degree of obtrusion. First, SC equipment is relatively cheap, especially since modern smart phones and tablets provide pocket-sized computers for real-time signal processing. Second, and very important, a small number of consumer SC devices can be used to capture data from several respondents at a time with minimal supervision, which allows for collection of data from a reasonable number of respondents in the short time frame allotted. Third, SC is also relatively unobtrusive, especially compared to consumer EEG, one of the methods which has also been employed by DR Audience Research in the past – EEG needs constant supervision as well as moistening of electrodes and the lab can only run one respondent at a time.

For the present study, we chose EDA as a measurement of arousal, since this is a common operationalization (see (Ravaja, 2004)). Within applied settings, there is arguably not yet any consensus on the methodology using EDA, and we see the present study as contributing to this discussion.

---

[3] Bartsch (2012) ties emotional engagement with characters to social and cognitive gratifications and thus to the more indirect emotional gratifications. This runs counter to most established cognitive theories of audience engagement in fiction, where character engagement is one of the key processes that lead to immediate emotional gratifications; the present study adopts this latter position.

[4] Deployment will typically select among these methods, e.g. focus groups and individual interviews are seldom combined.

4. Synthesized operational framework and research questions

4.1. Overall aim and assumptions

The overall aim was a research design that met the applied research needs of DR and was suitably informed by state of the art within applied media psychology and cognitive film theory. The design was based on the following assumptions:

1) Television content can be segmented into discrete *situations* (Tan, 1996). A prototypical situation involves characters situated in an environment with potential for conflict, as well as (implied) ways of resolving these conflicts.[5] Situations are thus designed for dramatic potential (Polti, 1921), by setting up, developing, and/or resolving conflicts between the main characters and their social and physical environment. These dramatic situations (described in screenwriting literature as consisting of smaller 'beats', in Blum, 2001; McKee, 1999; Newman, 2006), are realized as *scenes*. Our operational definition of scene is *a segment of the audio-visual material where one can observe unity of time and space as well as consistency of characters and environment.*[6]

2) The main driver of emotional engagement is the situations as experienced through one or more characters. Scenes can thus be categorized according to the characters involved and the possible emotional results; the sympathetic and antipathetic engagement with characters lead to emotional experiences that are either pleasant or unpleasant, depending upon viewer preferences. These emotional experiences involve arousal that can be measured in real time by way of EDA.

3) A given situation is both independent of and dependent upon other situations. A situation can be dramatic in itself – acts of violent crime or betrayals of trust are typical examples – but sequencing of situations can imbue a situation with additional dramatic force. A prior scene may lead into a particularly forceful act of betrayal by setting up the specific context, and acts of revenge necessarily depends upon prior wrongdoings etc. This dependency of character, situations, and scenes in sequence is captured by the term *storyline*. On such a view, the overall television drama sequence will typically contain several storylines tied to specific characters.

4.2. Research questions and hypotheses

Our two overall research questions were largely explorative but tied to the main interest of DR's audience studies as outlined above:

RQ1: To what extent would it be possible, based on SC measurements of viewer arousal, to identify moments of high and low arousal in a bottom-up fashion?

RQ2: In the case where one could answer RQ1 affirmatively, to what extent would it be possible to tie such moments of high and low arousal to the scene segmentation and to character concerns?

We were primarily interested in testing the overall research design and exploring how much information the purely quantitative element of the design would yield. We additionally assumed that qualitative data from focus groups could serve to triangulate and/or contextualize the quantitative findings.

Our theoretical literature enables a set of hypotheses concerning the measure of arousal, especially if one looks to literature on screenwriting. Almost all existing models of dramaturgy build on the Aristotelian notion of progression toward a climax (for examples see Blum, 2001; McKee, 1999 and Thompson, 2003). The simplest model would be that arousal rises linearly as the episode progresses, but most models of dramaturgical progression, including all modern Aristotelian versions, operate with a more complex structure, where the act structure of drama builds up with dips and peaks along the way. If this arousal-by-act model is correct, arousal should climb but also rise and fall visibly during the episode; this model is found in Tan and Blum, coming from cognitive film theory and screen writing respectively. The dominant models for film use three or four acts (see Thompson, 2003 and Newman, 2006 for television-oriented adoptions of this), but literature for television screen writing operates with up to seven acts (Blum, 2001). Based on this, we formed the following hypotheses:

H1: Arousal will ramp up in linear fashion, according to the simplest model of increased arousal as a function of dramaturgical structure.

H2: Arousal will ramp up in a pattern of ebbs and flows, according to accepted proposals of act-like structures within the screenwriting and academic literature.

H3: Arousal will be relatively constant, according to a market logic of grasping and sustaining attention.

As can be seen, these hypotheses are not mutually exclusive. We will now turn to a description of the research design and data collection procedures.

---

[5] This is arguably the dominant approach to segmenting fictional content, although "situation" is not always the term used. The typical segmentation unit in formal narratology is thus the event, and in many ways our concept of situation is compatible with the way the term event is used by such authors as Chatman (1978), Rimmon-Kenan (2002), and Abbott (2008). Our concept of situation, however, is more specific than merely referring to events, since it is intended to capture a set of constituent units as well as an event structure.

[6] If the characters and the locale stay the same and there are no large jumps in time, we treat the segment as one scene. Scenes may contain, so to speak, one or more situations but this discussion lies well outside the scope here.

5. Overall research design and methods

5.1. Recruitment of participants

The overall sample frame of participants was the members of DR Audience Research's own internet panel of approx. 9000 members. A total of 31 respondents were recruited, recruitment parameters being gender, age, educational level and a measure of appreciation of other DR drama series. Although the study was predominantly exploratory, we followed recommendations in the literature for confirmatory studies and set the number of respondents around 30 (as recommended by Kivikangas et al., 2011 and Field, 2012). The purpose of the sampling strategy was to get a heterogeneous sample that covered the major segments of Danish television viewers, i.e. a stratified purposive sampling. The initial sample consisted of 16 males and 15 females. Mean age was 43 years ($SD = 17$, median = 42, range = 22–69). Due to limitations of the SC equipment, the screening was carried out in groups of 4 to 10 persons. After screening, a short group interview followed where respondents discussed their experience with an interviewer. This qualitative part of the overall research design is an important and fairly extensive part of DR's overall evaluation procedure of series for in-house purposes, but in the present study the qualitative data was used mostly for triangulation and context; we bring in select results in the discussion following the quantitative findings.

5.2. Operationalization of skin conductance as a measure of arousal

As described in the general overview of psychophysiological research by Ravaja (2004) and in the more specific overview of EDA in Figner and Murphy (2011), skin conductance measurements are indirect measurements of the arousal of the sympathetic auto-nomic nervous system. This arousal accompanies various psychological processes and, as indicated in the previous section, such arousal is a key component of emotional processes, even if arousal can also indicate many other phenomena such as cognitive load and orienting responses. Skin conductance responses are additionally known to be relatively slow responses, with the typical delay assumed to be between 1 and 5 s (see Figner & Murphy, 2011) and Boucsein, 2012). Operationally, skin conductance can be measured by placing electrodes at the pointing and middle finger of the non-dominant hand and passing an electric charge between the two points. For the present study, 10 Mindfield eSense sensors were used together with 10 tablets (iPads) for data collection. The Mindfield application nominally samples at 10 HZ but the export function is only capable of exporting 5 values per second. The resulting intermediate data structure is one time series for each respondent, comprising a timestamp (5 Hz sampling rate) and corresponding measurements of skin conductance.[7]

5.3. Initial data processing and separation of phasic and tonic component

Skin conductance readings can be divided into tonic and phasic components (see Figner & Murphy, 2011). The most common measure of tonic components is the *skin conductance level* (SCL). SCL is not directly related to stimuli, but rather indicative of the general level of arousal. This measure describes the overall conductivity of the skin over longer time intervals, typically ranging from tens of seconds to tens of minutes. The phasic component, on the other hand, is measured by way of *skin conductance responses* (SCR). A single SCR is a discrete and short fluctuation in skin conductance that lasts several seconds and usually follows a characteristic pattern of an initial, relatively steep rise, a short peak, and then a relatively slower return to baseline. As an initial step, all time series were visually inspected for this distinct peak profile: Peak offsets are short and abrupt but decline slowly. If no such profile was verifiable the series was discarded, since the data could be invalid due to a noisy and unreliable reading.

The next step was a separation of tonic and phasic components of the signal. The main reason for running a tonic-phasic se-paration is to separate the event-related parts of the signal (i.e. the phasic component) from their moving baseline (i.e. the tonic component). In our data analysis, we focus on the phasic responses due to its relation to specific events and because it can be operationalized across shorter time intervals than SCL. Since our interest is in whether particular event structures (operationalized as situations) at particular times can be seen to contribute to arousal and engagement, the phasic component of the data is arguably the most relevant component for the present study. In order to separate the phasic and tonic components we used the Ledalab plugin for MATLAB authored by Benedek and Kaernbach (2010). The Ledalab plugin module offers several data manipulation procedures explicitly designed to handle EDA data series, and we used the continuous decomposition analysis function to separate the tonic and phasic components of our EDA measurements.

5.4. Normalization of the phasic component

Due to the nature of skin conductance, the magnitude of SCR readings can vary a lot from respondent to respondent; the in-dividual body's ability to conduct current and the level of sweat are the most obvious sources of variation. In order to be able to compare across individuals, the phasic measurements were normalised to vary between 0 and 1, i.e. a min-max normalization (Larose & Larose, 2014). The aim of this normalization is to control for between-subject variations in baseline and amplitude. One

---

[7] Electrodes were attached approximately 5 min prior to recording physiological data to ensure a stable electrical connection. The Mindfield eSense app records one person at a time and starting and stopping is done manually within the app: A test protocol was devised for proper synchronization between all respondents and the shown stimulus.

result of this normalization is thus that all measurements become more directly comparable and no single respondent's measurement can weigh significantly higher due to higher absolute skin conductance values (Latulipe, Carroll, & Lottridge, 2011); without any normalization, averaging procedures would be problematic.

Before conducting our actual data analysis, a total of eight series were discarded due to either no clear peak signatures or otherwise problematic data structures, e.g. artifacts from recording or data manipulation. We built a final dataset consisting of the remaining 23 time series measuring the phasic component and any reference to SCR in the following is to this dataset.

5.5. Identification of individual skin conductance responses

Using the dataset of phasic responses, we identified and counted peaks in every time series.[8] This calculation of peaks offers an alternative perspective on arousal: In contrast to average measures, the peak detection measures a single SCR above a certain magnitude and a peak thus indicates a discrete event giving rise to the skin conductance response. The distribution of such peaks gives an indicator of when the respondents responded to the events of the story without taking into account the magnitude of the response (apart from setting a threshold for irrelevantly small fluctuations in the data). As we will demonstrate below, both measures can yield information about possible emotional gratifications.

6. Stimulus material: Follow the money

6.1. Brief description of material

Our stimulus material was the first episode of a then new 10 episode television crime/melodrama series called *Follow The Money* (Danish title "Bedrag") in its "third-edit" version (which was very similar to, but not wholly identical with the final version broadcast). All episodes, including the first, have a running time of 58:30. *Follow the Money* depicts financial crime, and the series thus deals not only with morality, as is common in crime dramas, but also more specifically with greed and social injustice. The series depicts how financial speculation and crime impact people otherwise removed from the arena of the financial system (police, migrant workers, and local working class youth), thereby adding elements of melodrama to the crime plot. Described in one sentence, one might say that this is not a "*Whodunnit?*" but rather a "*How did they do it and what were the repercussions for ordinary folks?*". The first episode introduces three leading characters with individual storylines related to the overall crime plot centred on the company Energreen. Mads is a middle-aged and idealistic police detective with family issues investigating the death of a migrant worker at an Energreen project. Claudia is a young, ambitious female lawyer at Energreen who gets close, and perhaps dangerously so, to the executive level of the central firm. Nicky is a young car mechanic struggling to make ends meet as the breadwinner in his newly established family. Apart from those three central characters, the episode features the CEO of Energreen, Mads' wife and children, Claudia's immediate superior, Claudia's child and her ex-boyfriend, Nicky's girlfriend and his co-worker, a migrant worker, and two executives at Energreen.

6.2. Scene segmentation and descriptive statistics

Following our earlier definition of a scene, the first episode of *Follow the Money* (58.30 min) can be divided into 41 scenes each characterized by a key situation (or event) structured around one of the three main characters (we categorized three scenes as not belonging to any of these storylines). Our scene segmentation is included as Table 1. The overall mean scene-length is 83 s, and the median is 77 s. The maximum scene length is 250 s, minimum 25 s. We calculated a rough measure of screen time for each character in separate storylines by summing scene-lengths for the scenes belonging to each main character, as well as the mean and median scene-lengths. These descriptive statistics can be found in Table 2.

As Table 2 shows, Mads' storyline is the dominant in terms of screen time, followed by Claudia and Nicky. Mads' scenes tend to be longer, and the mean/median difference indicates that both Mads and Claudia get a few extraordinarily long scenes, which is confirmed by boxplots for scene-lengths by character, as visualized in Fig. 1. Fig. 2 visualises the individual character-centred scenes on the episode timeline and additionally maps actual scene length to the x-axis, i.e. the length of each block corresponds directly to screen time, which allows for some additional insight into the distribution of character-centred dramatic material. First, the two storylines of Mads and Claudia are distributed somewhat evenly across the episode, although there are two larger stretches in the middle part of the episode dominated by Mads (scenes 14–18) and Claudia (scenes 23–26) respectively. Second, there does not seem to be any clear pattern as to the placement of long and short scenes. Third, Nicky is left out of the last 11 character-centred scenes, which means that Mads and Claudia are the primary characters responsible for any arousal and emotional gratifications arising from character engagement in the last third of the episode.

---

[8] We used the R package "cardidates". The peak threshold was set to 2 times the standard deviation for each individual respondent. Although this package is not designed with SCR data in mind, visual inspection showed that it performed quite well in identifying the SCR peak structure.

Table 1
Scene segmentation of *Follow the Money* (episode 1).

Scene	Situation	Main character
1	A corpse is found at an Energreen workplace by the policeman MADS. The CEO of Energreen, is interviewed on television	MADS
2	INTRO	MONTAGE
3	MADS and children in kitchen with sclerosis-suffering wife in bed	MADS
4	MADS and children in car, worried	MADS
5	The young lawyer CLAUDIA is backstabbed by her immediate male superior at a meeting	CLAUDIA
6	MADS tells a migrant worker at an Energreen project that his son has been found dead	MADS
7	NICKY and girlfriend are out looking for an apartment	NICKY
8	NICKY and girlfriend quarrels about their situation	NICKY
9	NICKY meets young co-worker at the auto shop	NICKY
10	CLAUDIA talks to colleague about the backstabbing	CLAUDIA
11	CLAUDIA confronts her immediate superior about him backstabbing her but is ignored and tasked with finding an insider in Energreen	CLAUDIA
12	NICKY and his co-worker waits for an executive at Energreen	NICKY
13	The executive arrives. NICKY and co-worker are to fix a car for him. Afterwards the co-worker proposes NICKY that they steal luxury cars	NICKY
14	The migrant worker confirms the identity of his dead son	MADS
15	MADS interviews the migrant worker about the death of his son	MADS
16	Mads' wife reads to daughter and blames MADS for working too much	MADS
17	The CEO of Energreen is interviewed by MADS	MADS
18	MADS makes phone call about Energreen security	MADS
19	CLAUDIA calls her immediate superior about the insider	CLAUDIA
20	NICKY does not get the apartment due to high price and his co-worker repeats the previous proposal about stealing cars	NICKY
21	MADS interrogates the migrant worker about safety concerns	MADS
22	MADS and the migrant worker proposes to other migrant workers that they testify against Energreen	MADS
23	CLAUDIA finds a suspect in the insider case	CLAUDIA
24	CLAUDIA connects suspect to an executive (the guy with the car)	CLAUDIA
25	CLAUDIA confronts the suspect	CLAUDIA
26	CLAUDIA tells her superior that she has found the insider	CLAUDIA
27	NICKY gives the executive his car but is scolded severely	NICKY
28	NICKY and his co-worker arranges to steal the executives car	NICKY
29	MADS confronts his boss and is urged to take leave instead of pursuing the case	MADS
30	CLAUDIA'S ex-boyfriend delivers her son to her	CLAUDIA
31	MADS and his wife discusses whether he should take leave	MADS
32	MONTAGE	MONTAGE
33	CEO of Energreen is visited by a mysterious gentleman and they discuss the case	CEO of Energreen
34	CLAUDIA delivers evidence to her superior, and he tells the CEO of Energreen	CLAUDIA
35	MADS gets a call from the migrant worker, who has been fired	MADS
36	MADS visits the workers at Energreen and discovers that the migrant worker has committed suicide by hanging	MADS
37	Claudia's superior confronts CLAUDIA and then the CEO of Energreen, who promptly fires him	CLAUDIA
38	Mads' wife is suddenly better, family reunion and MADS cries	MADS
39	CLAUDIA prepares for a party and tells colleague that she has been invited to the CEOs home	CLAUDIA
40	CLAUDIA is promoted to head of legal counselling at Energreen	CLAUDIA
41	An investigator from internal affairs seeks out MADS and proposes that they take down the culprits at Energreen together. MADS obliges.	MADS

Table 2
Screen time and scene-length statistics.

Protagonist	Total screen time (Seconds and%)	Mean scene-length (seconds)	Median scene-length (seconds)
Mads	1640 (51%)	96	96
Claudia	1006 (31%)	77	61
Nicky	562 (17%)	70	70.5

7. Results

7.1. Phasic SCR measurements

We will start with visualizations of the aggregated skin conductance values (all references to average SCR in the following refer to the phasic component as described in the previous methods section). Fig. 3 shows a moving average of the indexed scores, using a moving window of 1 min (a more aggressive smoothing would leave out many of the fluctuations). Fig. 4 visualises the same data by plotting the average skin conductance level for each scene, still using the normalised phasic responses. As can be seen from comparing the moving average with the scene segmentation, the latter seems to capture most of the peaks visible in the moving average, although the scene segmentation seems to flatten some of the peaks a bit; this is expected since the mean scene length is a little above
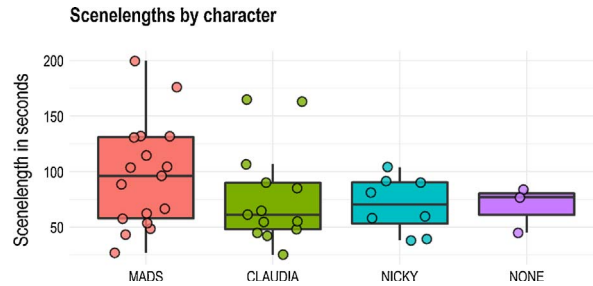
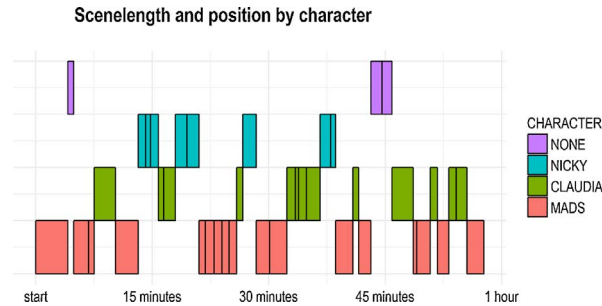**Fig. 1.** Box-plots with dots showing scene length by character.

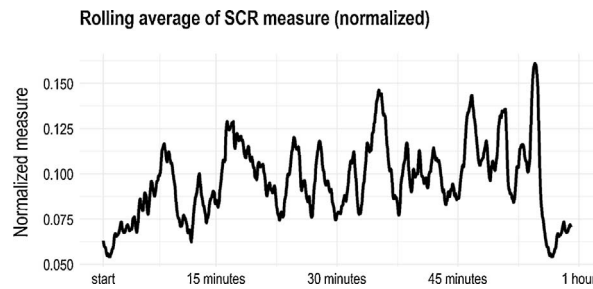**Fig. 2.** Scenes by character (storylines) across total episode.

Fig. 3. Rolling average of skin conductance measure (1 min window).
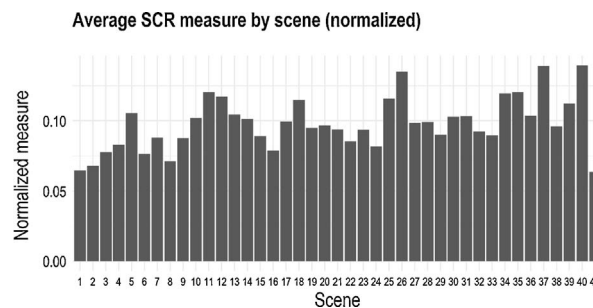

Fig. 4. Average skin conductance measure for each scene.

a minute (as described above). This could indicate that the timeframe of the scene structure may be too rough for identifying the true position of the peaks in aggregated responses. Since we did not segment our stimulus materials by individual events but by scenes, we are not able to contextualize individual peaks beyond their relationship with the situation structure inherent in the individual scenes. We do, however, revisit this issue below in section 7.4.[9]

Both the moving average and the scene segmentation indicate a slight upwards trend, and to test for a tendency across scenes, we conducted a one-way within subjects ANOVA with scene as the independent variable and average phasic SCR as the dependent

---

[9] Other factors speaking against a more precise segmentation are that our timeseries were only synchronized to within < 10 s as well as the inherent lag in EDA masurements.
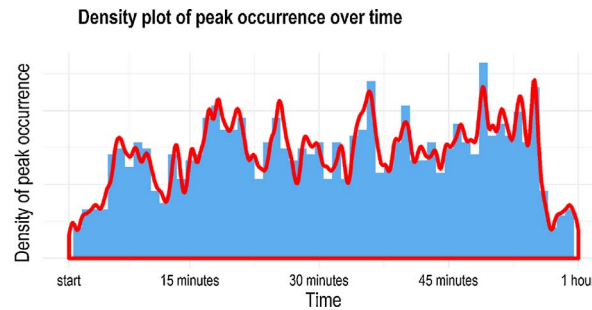
Density plot of peak occurrence over time

Fig. 5. Density plot of peak distribution across the episode.

variable (see Howell, 2009, 2010). Since the majority of the measurements by scene were right-skewed, the averages for each scene were log-transformed prior to the test. We were not able to test for sphericity using Mauchly's test (given our number of levels exceeded our number of respondents), but we assumed non-sphericity and corrected degrees of freedom using Greenhouse-Geisser estimates of spericity ($\varepsilon=0,24$). The overall test was significant (p = 0,001), and a test for a linear trend using an orthogonal linear contrast was significant (p = 0,008).[10]

7.2. Peak measurements

Turning to our measure of peaks, i.e. individual SCR events, Fig. 5 plots a density curve of peaks across the episode with every bin representing a one-minute interval − this corresponds roughly to the moving average in Fig. 3 and visualises where peaks occur by one minute intervals as the episode progresses. We see a similar pattern as in Fig. 3, with a slight but steady upwards trend. Fig. 6 plots the average number of peaks pr. second for each scene. This segmentation corresponds directly to that of Fig. 4 in that it segments by scenes and not regular intervals. We conducted a similar ANOVA as described above. Again, we were not able to test for sphericity using Mauchly's test (given our number of levels exceeded our number of respondents), but we assumed non-sphericity and corrected degrees of freedom using Greenhouse-Geisser estimates of spericity ($\varepsilon=0,343$). The overall test was significant (p = 0,025), and a test for a linear trend using a linear contrast was significant (p = 0,007).[11]

7.3. Comparison of SCR averages and peaks

Visual inspection of Figs. 4 and 6 indicate slight differences between average SCR and peak counts. To investigate this relationship, we conducted a Spearman's rank-order correlation to test correspondence between the two measures for the scene segmentation. The result was significant (rho = 0.79, p < 0.001) which indicates strong correspondence between these two measures.[12] Fig. 7 visualises both of these measures for the scene intervals: both have been min-max normalised to allow a visual inspection of the respective rank of scenes. The measures follow each other closely, but some of the scenes do differ in their relative rank (for instance 12, 18 and 35). We return to the possible implications of this in section 8.2. below.

7.4. Mean and median measures of SCR as indicators of arousing scenes

As described above, the scene segmentation may flatten the SCR average response somewhat due to scene length. To probe further for indications of particularly appealing scenes with the broadest appeal, we conducted two additional tests. First, we segmented all time series into 1-min windows[13] and calculated the mean SCR responses for each respondent within each window. To check the extent to which these arousal measures indicate overall agreement between viewers, we calculated the Intraclass Correlation (Hallgren, 2012; McGraw & Wong, 1996). We calculated a two-way, average-measures test for consistency with a resulting ICC = 0,649, which according to Cicchetti (1994) indicates a good level of consistency. We then adopted and simplified some of the procedures used in Fleureau, Guillotel, & Orlac (2013): We set the 10% quantile of the average measures of arousal as a baseline and conducted a series of one-sample t-tests and sign tests for each individual window to see whether the SCR levels differed from this baseline.[14] By selecting the sequences where both tests reached significance and tying these timecodes to the scene segmentation, a

---

[10] This ANOVA as well as the following leaves out scene 41, since this is part of the downwards tendency predicted by all dramaturgical models. With regards to the possible objection of assuming independence of measurements, the phasic/tonic separation ensures that all measurements reach baseline in between scenes. We do acknowledge that a procedure such as LME might be better suited to the data structure at hand.

[11] It should be noted here that the peaks data violate the assumption of normality rather severely.

[12] It should be kept in mind, however, that the correlated measures are both at the aggregate level.

[13] This is similar to the moving average in Fig. 1, except that the segments here are stationary, consecutive 60-s windows.

[14] The inclusion of the sign test (which tests against the median) should yield an improved indication of emotional congruence, as this guards somewhat against outliers; since peaks are normalised to between 0 and 1, one or two peaks can influence the average score disproportionately. Since this procedure involves 57 segments, we adjusted our alpha value using Bonferroni adjustment to alpha 0,05/57 = 0,00088. Bonferroni adjustment is generally seen as a conservative and the baseline used here is the 5th least arousing 60-s window in the episode, not a "no arousal at all" baseline.
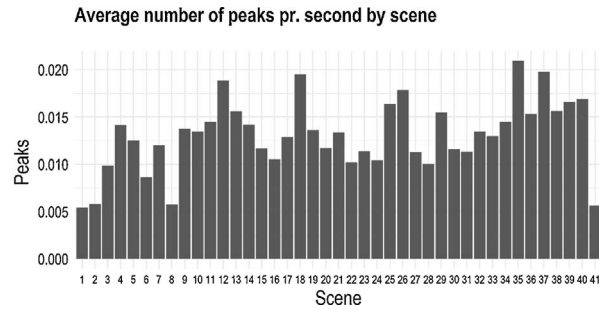
Average number of peaks pr. second by scene

Fig. 6. Average peaks pr. second by scene.

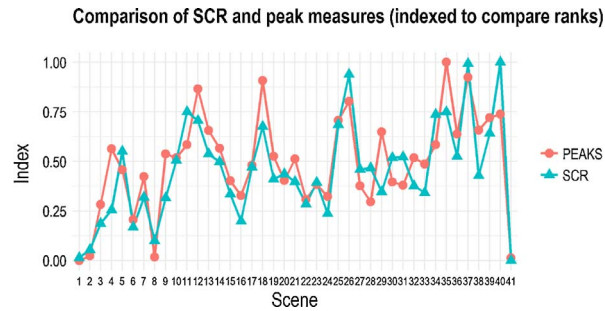**Comparison of SCR and peak measures (indexed to compare ranks)**

Fig. 7. Comparison of peaks vs. SCR for each scene.

series of scenes were identified as candidates for most collectively engaging. Both tests reached significance at p < 0,001 in three time windows, which correspond to four scenes, i.e. scenes 12, 21, 36 and 37. Scene 12 features Nicky and his mechanic colleague visit the "posh" Energreen offices. Scene 21, which is not a visible peak in any of the previous measures, features Mads interrogating the father of the deceased worker. The final time series window (around the 50-min mark) covers two scenes, namely 36 and 37 and these two scenes form the climax in Mads' and Claudia's storylines respectively. Mads' investigation into Energreen's neglect of safety climaxes in scene 36 with Mads discovering the corpse of the father (death by suicide) interviewed in scene 21. Claudia's investigation into insider fraud climaxes in scene 37 with her getting back at her previous boss for the aforementioned backstabbing and scene 39 develops this climax a bit further. Some of the scaffolding for this climax happens in scenes 25 and 26, and these were among the scenes closest to the significance level for both tests.

7.5. Summary of findings in relation to RQ and hypotheses

To sum up at this point, our two research questions cannot be answered conclusively from the present set of analyses. Any evidence is suggestive rather than definitive, and only correlational. Our first research question asked to what extent it would be possible to identify segments of high arousal in a bottom-up fashion and the procedure and findings described in Section 7.4. indicate that certain segments are more arousing that others. Our second research question asked to what extent these moments could be tied to our scene segmentation. This is obviously a crucial step for both applied and academic research into engagement with fiction, since we assume that any identified responses should correlate with, to some extent at least, the structure of the stimulus material. The findings here are suggestive rather than conclusive, but they do indicate that the two key storylines of Mads and Claudia work as intended, since the climaxes are among those scenes that show statistically significant increases in arousal.

We feel on somewhat firmer ground concerning the first of our hypotheses about arousal, namely that arousal would increase in a linear fashion. Visual inspection and statistical tests both indicate an aggregated audience effect where arousal ramps up slightly during the episode. Correlational analyses thus support the idea that viewers react to the dramaturgical development and climax of storylines with a linear increase in arousal. The second hypothesis was that arousal would ebb and flow in accordance with an act-structure model of dramaturgy, and this hypothesis gains some support from the visible peak-and-trough structure, although this is clearly more tentative; there seem to be at least four or five segments of engagement as indicated by the peaks and troughs across the episode. The third hypothesis was that arousal would be more or less evenly distributed across the episode, and this is to some extent supported by the results, since there seem to be no "dead periods" in the pattern of arousal.

The overall tendency is most clearly visible in Figs. 3, 5 and 7, where one can identify an initial baseline, a structure of peaks and troughs with a slight linear increase, and a final return to the initial baseline. All in all, this lends a modicum of empirical support to an entrenched trope within dramaturgical theory, namely the "excitement curves" that many dramaturgical manuals include, typically without any substantiation apart from the fact that "they just work". In the case of *Follow the Money*, the curve actually does work, although the relative increase in arousal is not exactly dramatic. We will now turn to a more general discussion of these findings.

8. Discussion

8.1. SCR, characters, arousal, emotion and appreciation

One of the goals of the study, as seen in our two RQs, was to tie emotional arousal to specific segments and, if possible, scenes and characters. This was only possible to a limited extent – climax scenes for Mads and Claudia are among the most arousing scenes, but the findings do not seem to allow for stronger conclusions beyond this. At this point, we can tie a connection back to the tri-partite framework outlined in the beginning and bring in the qualitative component of the study. The quantitative findings focus on immediate emotions measured as arousal, but it seems reasonable to expect that the two latter levels of the framework, which involve re-appraising immediate emotional experience, have some bearing on the overall viewer experience. Results from the qualitative interviewing can help bring these aspects to the fore. In the qualitative interviews after viewing, respondents were able to verbalize both a general positive response towards the show and its characters as well as an appreciation of the higher-order themes of the show, such as the nature of crime and social justice and family dilemmas and work-life balance. Several respondents found the episode exciting and identified Claudia's character as the main focal point of their enjoyment, although both Mads and Nicky were seen as interesting too. In fact, one result from the focus groups was that the mechanic Nicky and his friend were focal characters for several viewers, which may explain the arousal spike in scene 12. From our analysis of scenes, we did not expect this conversation scene to be particularly arousing, but it has an underlying element of social justice and features a character that many respondents turned out to be rooting for. With regards to the appeal of the two main storylines, the qualitative results were mixed. Mads was acknowledged as a main focus, but (in contrast to both Nicky and Claudia) the character was seen as lacking personal efficacy by several respondents. Mads' storyline is both more elaborate and dominant in terms of both screen time and scene length, but it pivots around a central motivational dilemma which blocks direct action. This dilemma makes Mads' storyline lean closer to melodrama than thriller and thus to less directly arousal-oriented emotions. In some contrast, Claudia was seen as efficacious, intelligent, and physically attractive, but also morally suspect. In general, Claudia's audience appeal may be explained with reference to both her physical attractiveness, intelligence, and her actions of deliberate risk-taking: This element of risk-taking connects her character directly to thriller-based fear-evoking situations and possible admiration of her courage under pressure, both high arousal concerns.

Besides the important issues of variance in responses between respondents hinted at in the comments about focal characters above, there is a more principled discussion here, which concerns the nature of different kinds of fiction-centred emotions. It is quite plausible that there are both short-term and longer-term emotional gratifications for viewers, which are not captured very well by real-time measurements of arousal only. For reasons of space, we will focus on the example of loss and grief and outline four points. First, scenes of death and loss can be interpreted as related to empathic sadness. Mild sadness is typically seen as an emotion with low arousal, and this may explain why scenes that deal with either mild sadness or loss may score low in real-time arousal. Second, Bartsch (2012) found that empathic sadness correlated positively with arousal, when both were measured by self-reports after viewing, and such scenes may thus be emotionally meaningful to viewers on a deeper level of later appreciation, i.e. the second and, especially, the third level in the tri-partite framework. Third, we would, on the face of it, expect responses to the loss of a child to be among the more universal of emotions, and this may explain why a scene such as scene 21 scores comparatively low on average but turns out to be significant in a test procedure focusing on median values. This type of situation may be low in arousal, but it has the potential to hit broadly and thus hook many of the viewers. A fourth relevant factor here may be the structure of sympathy touched upon in the first part of the article. Although the loss of a child is admittedly a very sad event, the migrant father is not a primary character: Rather, Mads is the primary gateway here, and this again may lead to tempered responses. As a final comment on the relationship between arousal, emotion and appreciation, it should be mentioned that the qualitative results point to appreciation of scenes and storylines that simply may not be high arousal material for any respondents (such as morning conversations with wife and children) but still quite important for overall appreciation. This type of scenes was mentioned as meaningful by respondents in the focus groups, and should thus not be ignored, as such scenes seem to be integral to storylines considered as meaningful on a deeper level.

8.2. Implications for future studies

We see our contribution as both theoretical and methodological. In addition to lending some support to a set of hypotheses concerning emotional engagement with television and the various content structures underlying such engagement, we hope to have contributed to the general toolbox of quantitative approaches used in applied and academic audience research. The present study indicates the viability of a range of similar research designs, which seek to connect measures of emotional gratifications to the content of audiovisual fiction. Our research design employed a stimulus segmentation into scenes and a measurement of interest and emotional engagement operationalized as arousal. Both of these aspects of the design can be generalized and developed further to produce timeline-based analyses of both audiovisual content and viewer reactions, either separately or in combination. In addition, we see the combined set-up as compatible with both aggregated and more case-based studies of emotional gratifications, where both the formal dramaturgical analysis and the SCR parts of the design may be used in conjunction with qualitative methods to probe individual and collective responses post-viewing, for instance through surveys, focus groups, and semi-structured or narrative interviews. We are ourselves working towards such research protocols.

Our study employed a method of stimulus segmentation into character-centred scenes and a measurement of interest and emotional engagement operationalized as arousal measured through skin conductance. Throughout the article, we have distinguished between a measure of arousal as an average of phasic skin conductance level and arousal as a measure of number of peaks indicating

discrete responses to events. We have taken these to be congruent measures of arousal (and the two measures were found to be highly correlated by way of statistical testing) but they arguably measure two different things. A high average SCR score indicates either a few highly arousing peaks or several lower peaks. A high average peak pr. seconds score indicates a high number of emotionally engaging events without taking into account the level of arousal. Put differently, average SC conflates frequency of events and level of arousal, whereas a peak count only measures frequency of events and disregards level of arousal. Future studies operationalizing SC as an indicator of engagement should take note of this relationship and take both of these measures and their specific indications into account. In the specific case of fictional content, the two measures in combination may be used, for instance, to identify scenes that are low on average arousal measures, but still manage to hook a sufficient number of viewers with emotionally engaging events, i.e. a relatively high number of low-to-medium sized peaks.

If future studies aim at uncovering general tendencies in emotional responses, the relationships that we have identified would probably be stronger with more homogenous samples. Although our sample size is reasonable according to the recommendations in the literature, the sample was deliberately heterogeneous since it also needed to work as part of a broader qualitative survey of responses across demographics. The demands of the applied setting may thus have worked directly against the typical aims of more deductive and confirmatory research. Further studies could look closer at individual cases as well as the possible aggregated audience effects across and between groups; in the latter case one might thus control for age, gender, social class, as typically seen in more deductive studies, and possibly bring in orientations toward fiction such as outlined by Suckfüll and Scharkow (2009). One might additionally consider statistical procedures such as LME, which might yield better control for individual differences.

Finally, we would expect future studies to benefit from using less complex stimulus materials, both in length and dramaturgic complexity. There are many variables in play in audio-visual material, and audio-visual fiction typically offers a continuous conflation of most, if not all, of the variables ordinarily described as efficacious within media psychology. Applied research on relevant material, i.e. actual television content, makes it very difficult to use highly controlled stimuli, but even within media psychology proper these variables are difficult to separate. We see at least two ways to proceed here. First, further work should be done on segmenting content structure, e.g. by way of cognitive film theory in combination with the type of qualitative content analysis seen in Kleemans, Eden, Daalmans, van Ommen, and Weijers (2017). Second, qualitative interviewing should be further integrated, since they seem better suited to uncovering emotional gratifications at the second and third level − especially the "slow burn" of fiction engagement. In any event, we see mixed methods as the most promising type of research design in future studies of emotional engagement with audiovisual fiction.

Acknowledgements

References

Abbott, H. P. (2008). *The cambridge introduction to narrative* (2nd ed.). Cambridge: Cambridge University Press.
Bartsch, A., & Viehoff, R. (2010). The use of media entertainment and emotional gratification. *Procedia − Social and Behavioral Sciences*, *5*, 2247–2255. http://dx.doi.org/10.1016/j.sbspro.2010.07.444.
Bartsch, A. (2012). Emotional gratification in entertainment experience. why viewers of movies and television series find it rewarding to experience emotions. *Media Psychology*, *15*(3), 267–302. http://dx.doi.org/10.1080/15213269.2012.693811.
Blum, R. A. (2001). *Television and screen writingFrom concept to contract* (4th ed.). Boston: Focal Press.
Bordwell, D. (1986). *Narration in the fiction film*. London: Routledge.
Bordwell, D. (2006). *The way Hollywood tells it. Story and style in modern movies*. Berkeley, CA: University of California Press.
Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). New York: Springer.
Carroll, N. (2003). *Film, emotion and genre engaging the moving image*. New Haven, CT: Yale University Press, 59–87.
Chatman, S. (1978). *Story and discourse. Narrative structure in fiction and film*. Ithaca, NY: Cornell University Press.
Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.
Field, A. (2012). *Discovering statistics using r*. London: SAGE Publications, Inc.
Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research: A critical review and user s guide*. New York, NY: Psychology Press.
Fleureau, J., Guillotel, P., & Orlac, I. (2013). Affective benchmarking of movies based on the physiological responses of a real audience. *Paper presented at the 2013 humaine association conference on affective computing and intelligent interaction* [2–5 Sept. 2013].
Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge: University Press.
Grodal, T. (1997). *Moving pictures. a new theory of film genres, feelings and cognition*. Oxford: Clarendon Press.
Gross, J. J. (1999). Emotion regulation: Past, present, future. *Cognition and Emotion*, *13*(5), 551–573. http://dx.doi.org/10.1080/026999399379186.
Hallgren, K. A. (2012). Computing inter-Rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.
Howell, D.C. (2009). Multiple Comparisons with Repeated Measures. Retrieved from https://www.uvm.edu/~dhowell/StatPages/More_Stuff/RepMeasMultComp/RepMeasMultComp.html#Howell .
Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont, CA: Wadsworth.
Kivikangas, J. M., Chanel, G., Cowley, B., Ekman, I., Salminen, M., Järvelä, S., et al. (2011). A review of the use of psychophysiological methods in game research. *Journal of Gaming & Virtual Worlds*, *3*(3), 181–199.
Kleemans, M., Eden, A., Daalmans, S., van Ommen, M., & Weijers, A. (2017). Explaining the role of character development in the evaluation of morally ambiguous characters in entertainment media. *Poetics*, *60*, 16–28. http://dx.doi.org/10.1016/j.poetic.2016.10.003.
Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data. an introduction to data mining* (2nd ed.). Hoboken, NJ: Wiley.
Latulipe, C., Carroll, E. A., & Lottridge, D. (2011). Love, hate, arousal and engagement: Exploring audience responses to performing arts. *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford: Oxford: University Press.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46.

McKee, R. (1999). *Story*. London: Methuen.

Newman, M. Z. (2006). From beats to arcs: Toward a poetics of television narrative. *The Velvet Light Trap − A Critical Journal of Film and Television*, 16–28.

Oliver, M. B., & Bartsch, A. (2010). Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research, 36*(1), 53–81 [10.1111/j. 1468-2958.2009.01368. x.].

Plantinga, C. (2009). *Moving viewers. american film and the spectator's experience*. Berkeley, CA: University of California Press.

Polti, G. (1921). *The thirty-Six dramatic situations*. Franklin, Ohio: James Knapp Reeve.

Ravaja, N. (2004). Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology, 6*(2), 193–235. http://dx.doi.org/10.1207/s1532785xmep0602_4.

Rimmon-Kenan, S. (2002). *Narrative fiction* (2 ed.). London: Routledge.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality & Social Psychology, 76*(5), 805–819.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172.

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition & Emotion, 23*(7), 1259–1283. http://dx.doi.org/10.1080/02699930902809375.

Silveira, F., Eriksson, B., Sheth, A., & Sheppard, A. (2013). Predicting audience responses to movie content from electro-dermal activity signals. *Paper presented at the Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*.

Smith, M. (1995). *Engaging characters: Fiction, emotion, and the cinema*. Oxford: Clarendon Press.

Suckfüll, M., & Scharkow, M. (2009). *Modes of Reception for Fictional Films Communications. 34* [pp. 361].

Tan, E. S. (1996). *Emotion and the structure of narrative film. film as an emotion machine*. Mahwah, N.J: Lawrence Erlbaum Associates.

Tan, E. S. (2008). Entertainment is emotion: The functional architecture of the entertainment experience. *Media Psychology, 11*(1), 28–51. http://dx.doi.org/10.1080/15213260701853161.

Thompson, K. (2003). *Storytelling in film and television*. Cambridge, MA: Harvard: University Press.

Vorderer, P., & Hartmann, T. (2009). Entertainment and enjoyment. In J. Bryant, & M. B. Oliver (Eds.), *Media effects. Advances in theory and research*(3rd ed.). New York: Routledge.

Wang, C., Wong, J., Xintong, Z., Röggla, T., Jansen, J., & Cesar, P. (2016). Quantifying audience experience in the wild: Heuristics for developing and deploying a biosensor infrastructure in theaters. *Paper presented at the 2016 eighth international conference on quality of multimedia experience (QoMEX)* [6–8 June 2016].

Statistics software , plugins, and packages used

MATLAB (R2010a)

IBM SPSS Statistics for Windows, Version 23.0

R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Arnold, J. B. (2017). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'.

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*(1), 80–91. http://dx.doi.org/10.1016/j.jneumeth.2010.04.028

Gamer, M., Lemon, J., &, I. F. P. S. (2012). irr: Various Coefficients of Interrater Reliability and Agreement.

Rolinski, S., Horn, H., Petzoldt, T., & Paul, L. (2007). cardidates. Identifying cardinal dates in phytoplankton time series to enable the analysis of long-term trends (pp. 997–1008).

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag

Wickham, H. (2017). tidyverse: Easily Install and Load 'Tidyverse' Packages.

Andreas Gregersen is Associate Professor at Dept. of Cognition, Media and Communication at University of Copenhagen where he coordinates the Cognition & Communication master's programme. His research focuses on video games, cognitive media theory, embodiment, and genre theory.

Birger Langkjær is Associate Professor in Film studies at Dept. of Cognition, Media and Communication at University of Copenhagen. His main research is in cognitive film theory, on sound and music in film and television, and on realism in Danish cinema. He has published three monographs as well as numerous articles and chapters in journals and edited volumes.

Lene Heiselberg is Senior Audience Researcher, Ph.D. at The Danish Broadcasting Corporation. Her research focuses on causes and consequences of emotional TV-programs and fiction series. In particular, she investigates how production of fiction series can benefit from pre-screening evaluations using a methodology of psychophysiological measurements and self-reports.

Jacob Lyng Wieland is Research Manager in BBC Audiences' measurement team working across cross media measurements, joint industry measurements, and data acquisition. Prior to joining the BBC he was innovation lead in The Danish Broadcasting Corporation's Audience Research team working on improving content testing by combining technology and the latest consumer neuro-science.