# Using Open-Source Tools to Measure Online Selective Exposure in Naturalistic Settings

Dominik J. Leiner, Sebastian Scherr, and Anne Bartsch

LMU Munich, Munich, Germany

**ABSTRACT**

Measuring audiences' selective exposure to media content in naturalistic settings constitutes a methodological challenge that has only partly been resolved. We present a new methodological approach that is based on the open-source web analytics software Piwik. This method allows for the tracking of selective exposure and facilitates the integration of selective exposure data with online survey data. To ease data handling, we created a plug-in turning Piwik into a scientific research tool. After discussing the theoretical and methodological background of collecting data on user selections, we provide step-by-step instructions on the integration of Piwik with online content, survey software, and the merging of tracking and survey data. Finally, we discuss research applications, advantages, and limitations of the new research tool.

Before media content can have effects, it first needs to reach its audience. This truism has become all the more salient with the proliferation of media channels and content choices that audiences have at their disposition in today's media landscape (Bryant & Davies, 2008; Hartmann, 2009; Knobloch-Westerwick, 2015a; Zillmann & Bryant, 1985). The freedom of choice that audiences have to select from countless sources of information and entertainment, as well as their increasing autonomy with regard to the timing, have spawned revived interest among media effects researchers in both micro- and macro-level effects of audiences' selective exposure to media content (Garrett, 2009b; Pariser, 2011).

In particular, new forms of access provided by online and mobile media have stimulated scholarly debate and research about the positive as well as problematic implications of audiences' increased autonomy in selecting their media diets. On the one hand, lay audiences' access to more diverse and previously inaccessible sources of information has been linked to democratic outcomes such as a well-informed public (Dahlgren, 2005) and audiences' active participation in public discourse (Bruns, 2005, 2008). On the other hand, individuals may also use their freedom of choice to avoid content that is inconsistent with their prior attitudes and interests. Consequently, selective media use may lead to greater political polarization for audiences who are caught up in attitude-consistent "echo chambers" (Jamieson & Cappella, 2008; Knobloch-Westerwick & Meng, 2011; Wallsten, 2005). Selective media use has also been linked to a deepening of knowledge gaps between well-informed audiences and those who engage in an "entertainment slalom" (Feldman, 2013; Feldman, Stroud, Bimber, & Wojcieszak, 2013; Knobloch, 2002; Prior, 2007) with a focus on apolitical entertainment media.

The recent revitalization of selective exposure research has also stimulated a number of significant methodological advancements (Clay, Barber, & Shook, 2013; Feldman et al., 2013; Hayes, 2013; Himelboim, Smith, & Shneiderman, 2013; Knobloch-Westerwick, 2015a). For example, research based on retrospective self-reports of selective exposure (Best, Chmielewski, & Krueger, 2005; Diab, 1979;

**CONTACT** Sebastian Scherr scherr@ifkw.lmu.de Department of Communication Studies and Media Research, LMU Munich, Oettingenstr. 67, Munich 80538, Germany.

Garrett, Carnahan, & Lynch, 2013; Sweeney & Gruber, 1984) or reported behavioral intentions (Fischer, Jonas, Frey, & Schulz-Hardt, 2005; Fischer et al., 2011; Garrett, 2009a) has been combined with observational methods (Bryant & Zillmann, 1984; Dillman Carpentier, Knobloch, & Zillmann, 2003; Knobloch-Westerwick, Sharma, Hansen, & Alter, 2005; Messing & Westwood, 2014; Zillmann, Hezel, & Medoff, 1980) as well as "big data" approaches that analyze selective exposure at the level of aggregated user data (Gentzkow & Shapiro, 2011; Webster & Ksiazek, 2012).

However, as Clay et al. (2013) have recently argued based on their analysis of the strengths and weaknesses of different methodological approaches, several methodological challenges remain that present nontrivial limitations to the validity of current selective exposure research. Among these challenges are the measurement of audiences' initial attitudes and their perceptions of the content, the continuous and repeated measurement of selective exposure to media content, and the study of selective exposure in diverse media contexts.

As discussed below, extant approaches (Knobloch-Westerwick, 2015a) have already overcome several of the methodological shortcomings, but they either require programming skills or costly access to commercial data sources and are not easy to implement for existing, third-party websites. In particular, the new methodological approach to the study of online selective exposure presented in this paper focuses on the latter aspect: A flexible implementation with different types of online content including both researcher-generated and third-party content—which is key to ensure high standards of external validity(Clay et al., 2013; Knobloch-Westerwick, 2015a). In an effort to enhance the current state of the art of selective exposure research, we developed a measurement tool that is based on open-source web analytics software and allows for a detailed and continuous observation of selective exposure to online content in the lab, as well as in naturalistic settings, including longitudinal designs. Moreover, it facilitates the integration of behavioral observation methods with online survey software—which is required to merge selective exposure data with introspective data from pre- and posttest questionnaires on participants' attitudes and perceptions of the content.

With these methodological desiderata in mind, we created a plug-in for the web analytics software Piwik, establishing a convenient interface for using the open-source software Piwik (download available from http://piwik.org/) as a scientific research tool. In this paper we provide step-by-step instructions for implementation by other researchers—in particular, concerning the integration of Piwik with websites and with online survey software (see Table 1 for a clarification of technical terms in this paper). We also provide instructions for the merging of selective exposure data with online survey data on a case-wise level. Finally, we discuss the specific contribution of our methodological approach to solving the challenges noted above, as well as its limitations, and implications for further research. For the convenience of the reader, the technical expertise required to understand the general data collection approach will be pretty low, the expertise required to follow the technical details of the software's operation then will be slightly higher, and the expertise required to actually deploy the solution is moderate.

## Theoretical and methodological background

The selective exposure paradigm is concerned with the fact that individuals' "voluntary exposure to information is highly selective" (Olson & Zanna, 1979, p. 1). The concept of selective exposure refers to situations where multiple media channels and/or content options are readily available to media users who tend to choose some of these options with greater likelihood than others (Knobloch-Westerwick, 2015a; Sears & Freedman, 1967; Zillmann & Bryant, 1985).

From a normative democratic perspective, audiences' freedom of choice in selecting their media diets is generally advantageous, because a well-informed public constitutes an important prerequisite for modern democracies (Dahlgren, 2005) and because the circulation of different positions and arguments is vital to the functionality of political discourse (Habermas, 1989). However, recent empirical research has also drawn attention to less than desirable consequences of multi-optional media environments that can impair political participation (Dilliplane, 2011; Knobloch-Westerwick & Johnson, 2014; Matthes,

**Table 1.** Explanation of the most relevant technologies and related expressions.

| | |
|---|---|
| database | Organized collection of data, based on database software that stores, organizes, and retrieves data. SQL databases can be imagined as a collection of spreadsheets (database tables). |
| database table | Organization unit in a SQL database, similar to a spreadsheet. A database table contains specific columns (variables/fields) and rows (entries/cases). |
| domain name | First part of an Internet (www) address (e.g., www.soscisurvey.com). The most common use of domain names is to address an Internet server (or network) in order to retrieve a web page from it. The Internet address's part after the domain then specifies which web page (or other resource) to retrieve. |
| fingerprinting | Method to map a large set of computer and browser characteristics (operating system/version, browser name/version, installed plug-ins, installed fonts, language setting, etc.) to a short string that, more or less accurately, identifies a single computer/user. |
| HTTPS | Internet communication protocol ("Hypertext Transfer Protocol Secure," based on HTTP) that encrypts the communication between the user's browser and the Internet server. Proper encryption ensures that third parties, who are involved in transferring the data (Internet providers, system operators, hackers, etc.), cannot read the contents of the communication. To prevent bypassing that encryption, current browsers usually disallow content from secure (HTTPS) and insecure (e.g., HTTP) sources to be displayed in the same browser window/tab. |
| IP address | Numeric address used by the Internet Protocol (IP) to identify a single Internet server or Internet user. Regarding the communication in the Internet, multiple devices may share the same IP address. Typically, one IP address is used per Internet access connection (e.g., per household or per company). The IP address of a household may vary over time. |
| JavaScript | JavaScript is a programming language often used within web pages. JavaScript runs in the Internet browser and performs tasks while the user (participant) is viewing a web page. Such tasks include modifications to the page content (e.g., display a menu) and transmission of data. |
| MySQL | Open-source software for SQL databases. |
| query parameter | An Internet address usually consists of four parts, of which the latter two are optional: The scheme (e.g., HTTPS), the domain name, the file/resource that is requested (e.g., index.html), and additional parameters added after a question mark (e.g., https://www.example.com/index.html?num=123) |
| SQL | Abbreviation for "Structured Query Language," a programming language that serves the communication with an SQL database. SQL databases are a widely used type of databases that organize data in a series of database tables. |
| subdomain | A domain name is resolved from the last element (the top-level domain, such as com, net, or de) backward. An Internet domain contains at least two elements (e.g., google.com) but may use another prefixed element, which is the subdomain, to distinguish between different services (e.g., mail.google.com and drive.google.com). |
| website vs. web page | A web page is a (single) document that is available via an Internet address (URL) and suitable to be displayed by an Internet browser. A website is a collection of related web pages, typically identified with a common domain name. |
| web/Internet server | The computer (may also be a virtual computer, a virtual service, or a cluster of computers) that is addressed when the Internet browser requests an Internet resource (e.g., a web page). The web server (or rather the server software) interprets the request and delivers a response (e.g., a web page) to the requesting Internet browser. |

2012; Mutz, 2002, 2006), increase political polarization (Garrett, 2009a, 2009b; Knobloch-Westerwick & Meng, 2011), or deepen knowledge gaps (Drew & Weaver, 2006; Prior, 2005). A meta-analysis of individual selective exposure to information (Hart et al., 2009) shows a moderate overall preference for congenial information ($d_{effect} = .36$; $N_{studies} = 300$) with slightly stronger effect sizes for political issues ($d = .46$; $n = 72$). These selective exposure effects were observed to be strongest for beliefs ($d = .53$; $n = 43$) and attitudes ($d = .42$; $n = 63$) as compared to behaviors ($d = .29$; $n = 194$). Beyond the context of political information, selective exposure can also take effect in domains such as business, legal decisions, consumer decision making, or interpersonal relations (Clay et al., 2013), which makes it a highly relevant paradigm not only for communication research.

Despite its undeniable merits, selective exposure research has also received methodological criticism early on (e.g., Frey & Wicklund, 1978, p. 138) and recently (Hastall & Knobloch-Westerwick, 2013). Nevertheless, it is widely agreed upon that even small selectivity effects can be relevant or that they may sum up to substantial influences (Clay et al., 2013). Therefore, it seems important to further advance the methodological repertoire of the selective exposure approach and to address the limitations associated with the current state of the art (Clay et al., 2013; Hayes, 2013; Knobloch-Westerwick, 2015a). Substantial technical hurdles, inherent in the nature of selective exposure research, need to be overcome to arrive at an unobtrusive and ecologically valid assessment of selective exposure behavior and to integrate such behavioral data with the introspective assessment of audiences' underlying motivations, attitudes, and perceptions (see Hastall & Knobloch-Westerwick, 2013; Zillmann et al., 1980). The following section provides an overview of the strengths and weaknesses associated with extant methodological approaches, with a special focus on observational measures of online selective exposure.

## *Methodological approaches to the study of selective exposure*

Research on selective exposure has covered a variety of media channels including radio, television, newspapers, and online media and has produced a considerable amount of methodological diversity (for recent reviews see Clay et al., 2013; Hastall & Knobloch-Westerwick, 2013; Hayes, 2013). Clay et al. (2013, pp. 148–149) critically reviewed existing techniques for measuring selective exposure and identified four broad methodological approaches: (1) studies using retrospective reports of individuals' selective media use, (2) measures of behavioral intention that focus on individuals' likelihood of selecting specific media content without measuring actual exposure, (3) observation of actual selective exposure behavior by tracking the quantity of consumed information (i.e., the number of sources and sometimes the time spent reading them), and (4) aggregate-level studies that use market data for describing selective exposure behavior on a population level. The respective strengths and weaknesses of these methodological approaches have been discussed by Clay et al. (2013) in detail and need not be reiterated here. Rather, we will focus on two key points that directly pertain to the focus of our methodological project.

First, an important limitation of introspective measures, such as retrospective reports and behavioral intentions, concerns their proneness to memory biases and socially desirable response tendencies (Clay et al., 2013, pp. 149–156). Still, introspective measures are indispensable, when it comes to the assessment of subjective background variables such as prior attitudes, perceptions of the selected content, or attitudinal changes that result from exposure to the content (Clay et al., 2013, pp. 163–164). Thus, the weaknesses of purely introspective approaches (1 and 2) can be a strength if combined with behavioral tracking approaches (3 and 4). To solve the conflicting demands associated with the assessment of different types of variables (selective exposure behavior vs. subjective attitudes), behavioral observation needs to be integrated with introspective measurement approaches (Hastall & Knobloch-Westerwick, 2013).

A second and related point, that follows from the need for integrated assessment of objective behavior and subjective attitudes, is that behavioral data are difficult to interpret on an aggregate level. Oftentimes, studies using aggregate data to investigate selective exposure do not assess preexisting attitudes toward selected (or selectively avoided) issues in the media, although this information is crucial within the selective exposure paradigm (Clay et al., 2013, p. 162). Moreover, in these studies it often remains unclear what elements of a message were de facto perceived by the study participants and actually influenced selectivity (Clay et al., 2013, pp. 160–163). The availability of individual-level data often is vital to the study of the antecedents and effects of selective exposure—as opposed to the study of selective exposure patterns per se.

Taken together, behavioral observation (i.e., tracking) of selective exposure to media content can be characterized as a methodological core element that should optimally be integrated at a case-wise level with the introspective assessment of subjective attitudes and perceptions (and additional variables that might be relevant for testing complex theoretical models such as possible mediator or moderator variables). The following section focuses on the current state of tracking-based approaches and considers their strengths and limitations.

### *Unobtrusive behavioral measurement of selective exposure: The current state of the art*

An early method for behavioral observation of selective exposure to television programs was developed by Zillmann et al. (1980), who presented their research participants with a choice of purported "TV channels" that were played back from video recorders and from which participants could choose using a "remote control" device that recorded exposure times to each of the programs. This method has successfully been employed in a number of laboratory studies (Bryant & Zillmann, 1984; Wakshlag, Reitz, & Zillmann, 1982; Zillmann & Bryant, 1985; Zillmann et al., 1980).

With the development of the Internet into a particularly powerful, multi-optional media environment, the focus of behavioral observation of selective exposure to media content has shifted to online content and computer or web-based research designs (Best et al., 2005; Garrett et al., 2013; Knobloch-Westerwick, Johnson, & Westerwick, 2015; Knobloch-Westerwick, Sharma et al., 2005; Messing & Westwood, 2014). Typically, studies investigate the impact of initial attitudes (and other individual-level variables measured before and after selective exposure) on the number of related news articles and time spent with reading them (Stroud & Muddiman, 2013; Taber & Lodge, 2006). Other studies investigated the influence of experimental manipulations on selection behavior as the dependent variable (Knobloch-Westerwick & Sarge, 2015; Valentino, Banks, Hutchings, & Davis, 2009).

A particularly prolific and elegant approach to observe online selective exposure has been developed by Hastall and Knobloch-Westerwick (2013). In a number of studies, their research group presented participants with specifically prepared websites that unobtrusively logged selective exposure measures such as the number of selected articles and the time spent reading selected articles within a given time frame, after which a follow-up questionnaire was automatically loaded (e.g., Knobloch-Westerwick, Dillman Carpentier, Blumhoff, & Nickel, 2005; Knobloch-Westerwick & Johnson, 2014; Knobloch-Westerwick et al., 2015; Knobloch-Westerwick & Meng, 2011). These studies comprise a wide topical array of issues (including gun control, abortion, health care, or minimum wages) and monitor participants' website browsing behavior (e.g., time spent with articles consistent or inconsistent with the individual's attitude).

Clay et al. (2013), who refer to this methodological approach as the "mock website paradigm" (p. 159), have noted several strengths and limitations (see also Hastall & Knobloch-Westerwick, 2013). Compared to other measures of selective exposure such as retrospective reports or behavioral intentions, this approach has the highest external validity as it unobtrusively captures actual selective exposure behavior. Hence, this method is less reliant on participants' accuracy of remembering their own past behavior or predicting likely future behavior, and it is less prone to attitudinal biases and issues of social desirability. Thus, as Clay et al. (2013, p. 159) have pointed out, "measures of observed selective exposure behavior have benefitted considerably from the advent of technology." These advantages include but are not limited to the monitoring of detailed information about the exposure duration, the number and quality of selected information, the order in which information was selected, and the overall diversity of information—which can be used for studying mediators or moderators of selective exposure effects (Clay et al., 2013, p. 159).

Clay et al.'s (2013) analysis of weaknesses associated with the "mock website paradigm" focuses on issues of generalizability associated with the restricted choice and a priori categorization of content included in artificial websites created for research purposes (typically including 6–12 articles, see Hastall & Knobloch-Westerwick, 2013, p. 103). Selective exposure effects may differ if—rather than focusing on a (forced) selectivity task within a given sample of news articles—participants have the choice of selecting completely irrelevant content (Feldman et al., 2013) or when nonpolitical content is available (Arceneaux & Johnson, 2013). Thus, despite its higher level of external validity compared to retrospective reports and behavioral intentions, this approach may only partly reflect real selective exposure behavior outside the lab. Moreover, Clay et al. (2013) argue that researchers' preselection of articles involves untested a priori assumptions about participants' interpretation of the content.

We think that both of Clay et al.'s (2013) criticisms can (partly) be addressed within the mock website paradigm, for instance, by including irrelevant content in addition to the target articles and by carefully pretesting and adding manipulation checks of how participants perceive the stimulus materials (see Hastall & Knobloch-Westerwick, 2013). However, as Hastall and Knobloch-Westerwick (2013) have noted, "web server–based measurement of selective exposure to online content typically requires a substantial technical knowledge of Internet technologies and protocols" (p. 103). They conclude that "potential future enhancements, which we welcome, could include the development of open-source software tools for convenient implementation in different research contexts or the development of easy-to-use templates implemented in existing software tools" (Hastall & Knobloch-Westerwick, 2013, p. 104).

The present methodological development aims to address existing technical challenges in an attempt to make the tracking approach free of charge and accessible for researchers without extensive technical background knowledge. Moreover, we aim to advance the integration of technologies for tracking selective exposure with preexisting websites and with online survey software.

## Using Piwik for unobtrusive measurement of selective exposure to online content

Piwik is a software package for web analytics, collecting and analyzing data on the use and users of a given website (Karg & Thomsen, 2011). Its statistics may include information about search terms, website referrers (the website that users have seen before), time spent on a website and on specific web pages, a website's typical landing and exit pages, IP addresses, web browsers, etc. Piwik is open-source software (GNU Public Licence v3) programmed in PHP, that is provided at no cost, has an active community, and is widely used around the globe. It requires a MySQL database running on the web server as storage for the monitored data. Of particular interest with regard to privacy and research ethics, Piwik can be run on a local server, i.e., as "in-house-solution" (Karg & Thomsen, 2011, p. 489). That offers possibilities to ensure highest data security standards, handling of user data is possible without involvement of third parties, and data need not necessarily be transferred to foreign countries. Compared to other web analytics tools such as Google Analytics, for example, a local Piwik server further allows the researcher full access to *all* detail information stored in the MySQL database via database interface or the structured query language (SQL).

Using the web analytics software Piwik, any given web content, including preexisting online content, may be employed for selective exposure research. This and the simple use of Piwik are important advantages and prerequisite for creating study designs with high standards of external validity. To implement accurate user tracking, any web analytics software requires some modification of the original web content. In case of Piwik, a piece of JavaScript code needs to be added to the source code of the target website; if the website is based on a common content management system (CMS) instead of static HTML, ready-made plug-ins are available to implement the necessary JavaScript. Thus, researchers can either track selective exposure to self-published websites or conduct studies in cooperation with other content providers such as news media, companies, and other organizations who are willing to implement the tracking script on their website. This possibility of tracking research participants' selective exposure to preexisting online content constitutes an important complement to research using mock websites designed for research purposes (Hastall & Knobloch-Westerwick, 2013). For instance, unobtrusive measures of selective exposure to online content could improve our understanding of online news selectivity (see Knobloch-Westerwick & Meng, 2009; Messing & Westwood, 2014) or could contribute to selective political news reading and sharing (e.g., Kang, Lee, You, & Lee, 2013).

A previously unresolved challenge of using web analytics software such as Piwik for research purposes is the lack of data export functions on a case-by-case basis, which is required for case-wise data analysis and merging of selective exposure data with survey responses. Highly automated merging of background survey responses that assess participants' attitudes before and/or after selective exposure to media content, for instance, allows large-scale online studies that yield sufficient statistical power to work with the complexity of realistic websites. To address the challenge of data extraction, we created a specialized plug-in for Piwik that offers convenient data export functions including case IDs for individual research participants to merge tracking and survey data stemming from standard survey software such as SurveyMonkey, Lime Survey, or SoSci Survey.

## Technical background and practical application

Exposure of study participants to online content and selection of content typically involves an Internet browser. Every click on a hyperlink, button, or advertising item while surfing a website is an active selection of content. These are machine-readable interactions that constitute valuable information for researchers, who may use this information to observe individuals' choices in an unobtrusive way. The methodological challenge is to (a) record these choices while distinguishing different participants and (b) extract those selections that are relevant for the specific research design. Whenever the user clicks on a hyperlink to see the content behind, the request is an action that is by default restored into a "logfile" (mostly with further information such as time and the IP address of the requesting computer). These logfiles are usually available on the web server that handles the requests and can generally be used for monitoring Internet behaviors. If the server is not run by the researcher, a proxy server may be employed in laboratory settings (for this alternative monitoring approach see Menchen-Trevino & Karr, 2012). While the information provided by logfiles is useful under some conditions, it suffers two major shortcomings. First, an IP address does not always identify one single device (computer, tablet, cellular phone, etc.). One device may use multiple IP addresses, or several devices may share a single IP address to access the web server. Second, browsers may store some content in their cache (not requesting it from the server, if selected again) or automatically prefetch web pages that the user is likely to select next.

Overcoming the first shortcoming is known as user tracking. This means to accurately tell which device a request was sent from—based on the assumption that a device is used by only one person at a time. The most widespread method for user tracking is a cookie. A tracking-cookie is an ID code that is stored on the user's computer and sent along with every request to the web server. More recent methods of user tracking include fingerprinting, which becomes relevant if the browser does not actively support the tracking by sending a cookie. To overcome the second shortcoming, even more cooperation from the user's Internet browser is necessary: It must actively send information to the observer when a page (or other content) is displayed on the user's screen. Currently, JavaScript is the most common and widespread technique for such active content. An appropriate script could send a notification when a page loads, when a specific item comes into sight, when a button is pressed, or when the user switches between browser tabs. As Internet browsers send such notifications in the background, a user does usually not take notice of them.

## Using Piwik for recording users' online activities

Piwik is attractive to researchers because it is free and it works immediately with minimal customization. In the default configuration, Piwik employs cookies and JavaScript to track website users, ensuring the distinct identification of participants throughout a website's pages. If cookies are blocked or unavailable for other reasons, Piwik will fall back to fingerprinting. The JavaScript, which must be included in every web page (see below), signals the Piwik server whenever a new page is displayed in the participant's browser. This mechanism avoids issues with browser-side caching or prefetching. According to Hastall and Knobloch-Westerwick (2013, p. 96), valid, unobtrusive monitoring of selective exposure must ensure that (1) each accessed web content is recorded, (2) no exposure record exists for content that has not been loaded or seen, and (3) the time spent with each content allows an exact reconstruction of individual online selectivity. All of these criteria are met when using Piwik for monitoring web activity. Specifically, Piwik is able to track every browser activity, including the (repeated) use of browser back and forward buttons, and monitor all kinds of opening a new webpage by the participants while tracking the respective time for each action in seconds. Hence, issues with eventually locally cached and retrieved copies of websites, prefetching, or "privacy features" activated in browsers do not apply for Piwik.

As mentioned above, Piwik is only one possible solution to collect data on selective exposure. Apart from logfile analyses, there are plug-ins (e.g., Web Historian for Google Chrome) and other local software to collect and transmit use data. One important advantage of Piwik is that its application does not depend on client-side software, i.e., the participant neither must install a program/plug-in nor actively send data. Piwik also ensures adequate privacy by not collecting data on Internet surfing aside the stimulus (see http://piwik.org/docs/privacy/).

To ensure highest standards of privacy, we recommend using a local web server for both Piwik and the stimulus content (the server requirements are PHP and MySQL) instead of cloud services. Moreover, there are three steps for the researcher to use Piwik to record selections on a website: (1) Make the stimulus material available through the Internet or the local network, so that research participants can access the website (via HTTP or HTTPS). The stimulus material may be located on the same webserver as Piwik or not. Most important, both Piwik and the stimulus should be available through the same protocol (HTTP or HTTPS) in order to comply with browser security principles. In contrast to other proxy systems for monitoring (Menchen-Trevino & Karr, 2012), our tool allows monitoring websites with HTTPS security. (2) Set up Piwik on the web server and, optionally, activate our plug-in ExposureResearchTools to conveniently download the recorded data. (3) Add a piece of JavaScript to the stimulus material.

(1) **Make the stimulus material available through the Internet or the local network, so that research participants can access the website (via HTTP or HTTPS)**. The stimulus material may have been created as static website (HTML files) or by means of a content management system (CMS) such as WordPress or TYPO3. We recommend a static website, because it requires no CMS setup, integration of JavaScript code is straightforward (copy and paste to HTML file), and distinction of pages based on their URL is trivial.

(2) **Set up Piwik on the web server and, optionally, activate our plug-in ExposureResearchTools to conveniently download the recorded data.** The setup of Piwik on a web server is explained in a detailed installation manual on the Piwik website, including a video tutorial. If different stimulus websites share the same Internet (sub-) domain, it is not necessary to create (register) multiple websites in Piwik, as the websites and their pages will be distinguished by their path names. Yet, if different websites are registered, the ExposureResearchTool will create separate data files for each.[1]

(3) **Add a piece of JavaScript to the stimulus material.** The JavaScript code required for user tracking is available from the Piwik user interface (already displayed during the setup process; see Figure 1) within the administration menu where the monitored websites are listed. If using static websites, simply "copy & paste" this code into every HTML file (e.g., immediately before the closing </head> HTML tag, which should already exist in the HTML file). If multiple websites have been defined in Piwik, please note that the JavaScript code varies per website (but not per web page). If necessary, upload the modified files to the web server. If employing a CMS-based website, the Piwik manual includes detailed information on how to add appropriate JavaScript. When the stimulus material is viewed with an Internet browser, Piwik will now record every page view and store these log data (who opened which page at what time) in a MySQL database.

## *Combining Piwik with survey or observational data*

After having introduced the basic principles of preparing Piwik to measure online activities on a given website, we will now explain how to realize the redirect from a survey to the stimulus material in a new browser tab. The examples are based on a questionnaire created with SoSci Survey, which is a software for

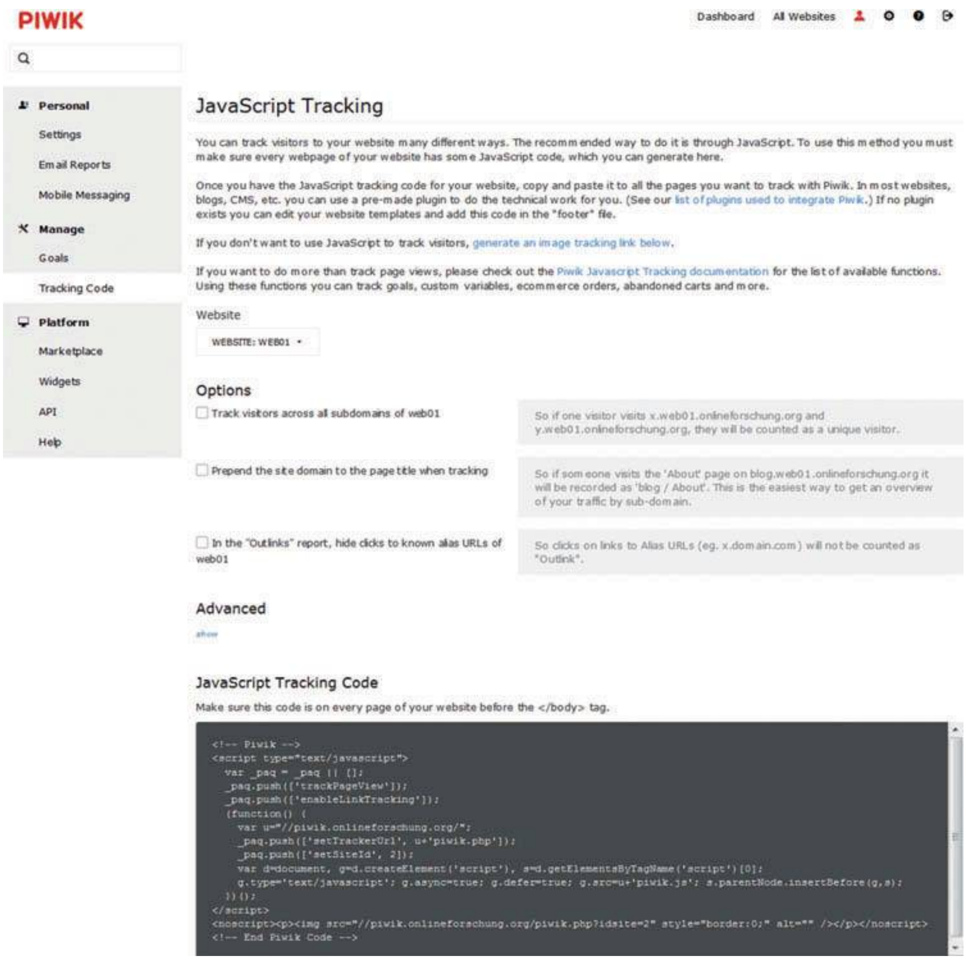


**Figure 1.** Example of JavaScript tracking code.

conducting online surveys that also supports enhanced research designs. As a first example of a typical study design using Piwik, we describe how this tool could be implemented to replicate and extend research conducted by Knobloch-Westerwick (Knobloch-Westerwick, 2015a, 2015b) to test the Selective Exposure for Self- and Affect-Management (SESAM) Model. This research design involves collection of survey data through an online questionnaire in a pre-/posttest design with selective exposure as a mediating variable. After answering questions about predispositions or preexisting attitudes, participants are redirected to a manipulated website containing news or entertaining content. Media selectivity is then unobtrusively monitored using Piwik during a given time frame, in which a new browser tab is open. After that, attitude change and/or behavioral intentions are assessed in a post–selective exposure questionnaire to which we redirected participants automatically. This research design allows researchers to further examine the SESAM Model (Knobloch-Westerwick, 2015a, 2015b). Knobloch-Westerwick (2015a, pp. 375–381) used a comparable research design to empirically test the model in different contexts (gender, race and self-esteem, political communication, health communication), especially with regard to the assumed underlying dynamics of the model (e.g., the influence of repeated selective exposure to media content); further research using the flexibility afforded by Piwik in terms of examining exposure to preexisting websites could make important contributions to our understanding self and affect management in naturalistic settings.

To combine survey data with Piwik monitoring data, the stimulus' URL (http://www.example.com/stim/s01.html) has to be extended by a query parameter (i.e., a GET variable containing the participant/interview ID, such as http://www.example.com/stim/s01.html?num=1). Piwik will store the complete URL of any page visit—including the query parameter "num," which identifies the participant in the Piwik records and can be extracted from the data later.

To open the redirect URL in a new tab or pop-up window using a hyperlink, HTML code is added to a questionnaire page (see Figure 2). The survey software must add the current participant's ID to the hyperlink (SoSci Survey will automatically replace the placeholder *%caseNum%*, for example). Depending on the research design, it can be helpful to use a hyperlink or JavaScript to open a new tab (or window)—depending on whether the participants can freely browse the media content or whether they were given a limited time frame for exposure to media content. For the convenience of the reader, we provide a JavaScript example along with the Piwik plug-in ExposureResearchTools that will be described below.

### *Example study procedure using Piwik to full advantage*

As Piwik can be employed freely by researchers for capturing online selectivity for both preexisting and mock websites, we think it can make an important contribution to selective exposure research. With other applications for selective exposure research in mind, all of which have specific strengths for answering specific research questions, we want to outline a recent study employing a research design that specifically benefits from the capabilities of Piwik as a research tool. The study (Kretzschmar & Waßink, 2016) was conducted by the Universität der Bundeswehr München and the LMU Munich in cooperation with a governmental department (Federal Office of Civil Protection and Disaster Assistance, FOCPDA) and focused on the impact of risk awareness videos on subsequent information-seeking behavior on the FOCPDA website. To implement Piwik on the third-party website of the government, their website was mirrored and transferred to a university web server. Study participants were then directed to the mirrored version of the website that looked just like the de facto FOCPDA website, except for unobtrusive changes within the URL. By doing so, study participants could be tracked without monitoring regular visitors of the governmental website. An invitation to participate in the study including a hyperlink to the online questionnaire was sent out via email. The starting page of the survey included an informed consent form that browsing behavior will be monitored. The online survey included a professional stimulus video on natural disaster protection that had been created for the FOCPDA website. After watching the stimulus video and answering questions (e.g., preexisting attitudes and personality traits), people were directed to the mirrored website and were told to freely browse the website for a limited amount of time. After browsing the website, participants were redirected to the post-questionnaire to evaluate the website and to provide answers on topic-related attitudes, behavioral intentions, and demographics. This research design allows not only to explore the influence of individual-level variables (such as preexisting attitudes) on browsing behavior, but also short-term influences of an audiovisual stimulus addressing a related topic on subsequent selective information seeking. Moreover, using this kind of research design, selectivity can be modeled as a mediator between individual-level pre- and post-exposure attitudinal measures and can be linked with the effects of the experimental stimulus—in sum, the design opens new pathways for refining our knowledge about selective exposure to media content. Piwik allows for implementation on existing third-party websites (or mirrored versions of it) due to an open interface and common implementation code (as described

```
<a href="http://www.example.com/stim/s01.html?num=%caseNumber%"
target="_blank">View Website</a>
```

**Figure 2.** Sample HTML link used to redirect participants from SoSci Survey to Piwik-enabled stimulus website.

above), which offers new opportunities for externally valid selective exposure research. As the project in collaboration with the FOCPDA is ongoing, we cannot present data here. Nevertheless, this exemplifies that privacy settings are high enough to conduct web tracking studies of selective media exposure in collaboration with government authorities.

## *Extraction of data using the ExposureResearchTools*

Finally, we describe how to obtain and prepare the data gathered with Piwik. Piwik records all browsing activities in a MySQL database. For every website visit (a series of views of related webpages, or in case of a study, one single participation), Piwik creates a unique entry and ID and stores all browsing information related to this case ID. With the information, it can be reconstructed at what time (measures in seconds) a specific page was viewed, for how long, and in what order during the visit (including back and forth movements within the web magazine).

For the convenience of researchers who use statistical software and to simplify data retrieval from the MySQL data set into a ready-to-use data set, we developed a plug-in for Piwik called ExposureResearchTools, available from the Piwik Marketplace (https://plugins.piwik.org/) or upon request from the authors. The plug-in adds a menu item "Research Tools → Export Visits." This feature provides a dialogue with several options (see Figure 3) to customize the merging and export of data from Piwik's MySQL database into a data file. The output format CSV (character-separated values) is compatible with SPSS, Stata, GNU R, and other statistical and spreadsheet software. The "Export Visits" dialogue includes options for selecting one of the websites that have been monitored, (optionally) the variable from which to extract the participant ID, and for refining the data-preprocessing and structure. If the option "Skip visits without subject ID" is selected, only those cases of research participants are extracted that have been redirected from the background survey to the website. The variable specified under "Read subject ID from GET variable" is included as "CASE"

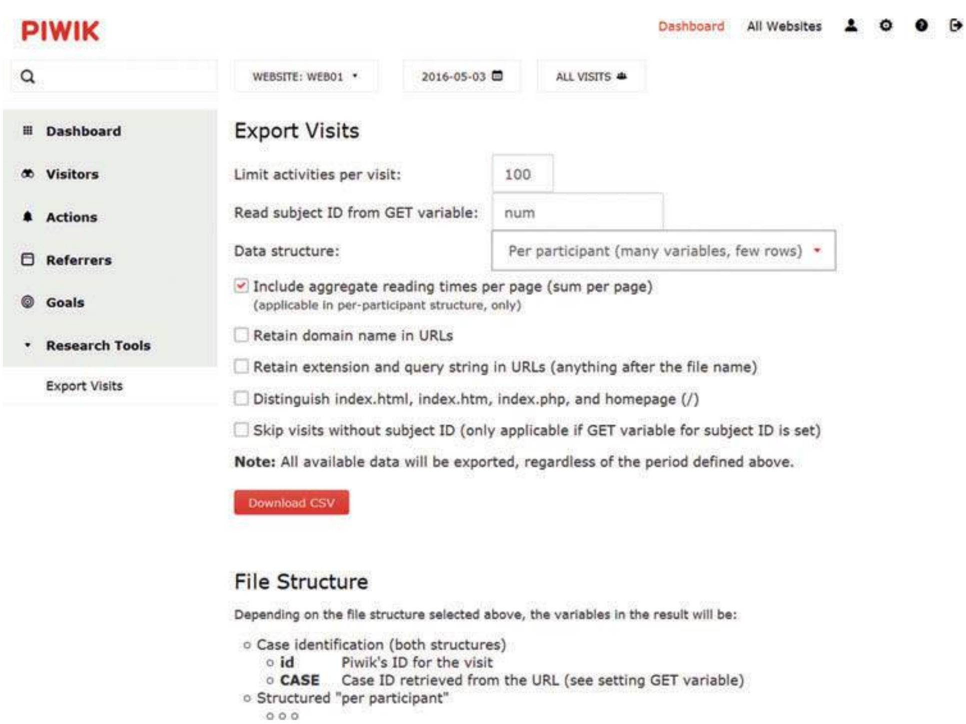


**Figure 3.** Piwik plug-in "ExposureResearchTools" for downloading monitored data from the SQL database.

in the exported CSV data file to facilitate the merging of tracking data recorded by Piwik with data from an online survey. The number of activities (web page views or clicks) per visit and the time spent for each activity (in seconds) is included in the export data file and can be limited to avoid exceedingly large data sets (the default limit is set to 100 activities). Usually, these default values should be fine with most online or laboratory studies in which participants are instructed to browse within a specified time span of 4 to 5 minutes (Hastall & Knobloch-Westerwick, 2013). In addition, researchers can make several choices that influence the structure of the exported data. Researchers may choose between aggregate data per participant (i.e., order of visited web pages and, optionally, aggregate reading time per page) or per page view (i.e., a data set suitable for multilevel analyses, providing one record/row per activity). By default, the tool removes parts of the URLs that are usually irrelevant for selective exposure analyses, but a series of "retain" options is available to keep these parts.

The dialog also includes a description of the variable labels in the exported file. We used intuitive labels for monitored data, such as A1–A*n* for performed activities (i.e., pages viewed during the visit), T1–T*n* indicating the time spent for each activity (i.e., T1 is the time spent for A1 in seconds and so on), and AT_*xyz*, indicating the aggregate time spent per activity (i.e., the total time spent on one particular page).

For capturing the last reading time on the website where selective exposure is measured, there is a special option[2]: We included the variable "AT," indicating the aggregate time of all monitored actions on the website (excluding the time spent with the last page visit, which is not recorded). The difference of the selective exposure duration and the duration stored in "AT" is a good estimate for the time spent on the most recent page.

### *Handling and merging exported data using SPSS*

After exporting and downloading the CSV file from Piwik, it can be imported in SPSS or any other statistical analysis software that automatically converts it into a data set. The ExposureResearchTools offer two different ways of data preparation: Data are either in a sequential case-wise format that describes the sequence of visited websites per participant, which can be merged with data from background surveys on the same individual, or data are prepared in a hierarchical format for multilevel analysis that describes every single action on the tracked website a single case.

The Piwik data can be merged with survey data using the SPSS "merge files" function in SPSS under the "data" dialogue by adding new variables (from the Piwik data file) to the survey data file. To match cases on a key variable, the participant ID (variable name "CASE" in the Piwik data file) needs to be matched with the variable name in the survey data set. Thus, SPSS will add the Piwik data as new variables to the survey data set at the end of the variable list so that both survey data and tracking data from Piwik refer to the same study participant's individual case ID.

## Discussion

This paper presents a new approach to capture unobtrusive measures of selective media exposure that is fully integrable with custom online survey tools such as SurveyMonkey, LimeSurvey, or SoSci Survey and that offers the possibility to model selective exposure as either an dependent or independent variable of media effects within pre-/posttest research designs. Based on their analysis of the current status of the selective exposure paradigm, Clay et al. (2013) concluded that "several overarching methodological issues limit the ability to draw consistent and meaningful conclusions from prior work" (p. 163). As discussed above, these include (1) the specificity of measurement of initial attitudes, (2) the continuous measurement of selective exposure measures to media content, (3) the measurement of audiences' perception of the content, (4) the study of selective exposure in diverse media contexts, and (5) the implementation of longitudinal designs to examine the effects of repeated exposure to specific sources. We aimed to provide researchers with easy-to-implement solutions to overcome challenges 1, 2, 3, and 5 based on the integration of open-source web analytics software with regular websites and online survey software. Specifically, we developed a plug-in for Piwik, a free-to-use web analytics tool, to measure online behaviors on any given website.

With regard to challenge 2, Piwik offers flexible procedures for measuring selective exposure to online content. Researchers can monitor and track all of the users' online selective behaviors and online activity, such as clicks on articles, the time spent on each website, and the order in which the content was selected. There are other monitoring techniques that yield similarly unobtrusive measures of exposure to relevant media content. For instance, Bakshy, Messing, and Adamic (2015) used Facebook's log data to explore people's selective exposure to crosscutting political content posted by their friends on Facebook. While most researchers probably will not have access to such data, we have good news for all those who are interested in the underlying mechanisms of homophily-influenced selective exposure to socially recommended/shared crosscutting political media content: Piwik's strong suit (besides being accessible to everybody) is its ability to track the time people spent with such political media content. Using news websites programmed by researchers to present news in a context suggesting that the news has been liked by other people sharing the same interests or political views, Piwik would allow researchers to disentangle how long people are actually reading news article teasers before selecting an article for more in-depth reading. The default monitoring process provides researchers with raw data on this, which can be differently aggregated using our Piwik plug-in ExposureResearchTools. There is an option for downloading sequential, individual-case format data, and another option for downloading data in a hierarchical format for multilevel analysis. The former option may also include aggregate exposure times for each of the web pages being tracked, which constitutes the most widely used continuous measure of selective exposure (Hastall & Knobloch-Westerwick, 2013).

In terms of challenges (1) and (3), the integration of Piwik with online survey software allows for a case-wise matching of behavioral measures of selective exposure with a differentiated introspective assessment of individuals' prior attitudes and with their perceptions of the content using pre- and posttest surveys. In addition to these challenges noted by Clay et al. (2013), the integration of behavioral data with survey data also allows for the assessment of moderating variables (e.g., personality traits) or mediating variables (e.g., cognitive and affective responses to the content selected). Thus, the implementation of study designs for testing complex theoretical models based on both behavioral and introspective data is made accessible to researchers without extensive technical background knowledge. The Piwik plug-in provides a tool that helps convert data from Piwik's MySQL database into a CSV file that can be conveniently handled by most statistical software packages and that includes participant IDs for case-wise matching of tracking and survey data.

With regard to challenge (5), the integration of Piwik with actual websites and with online survey software such as SoSci Survey allows for longitudinal designs. Participants of an online access panel can be invited at several points in time to visit a given website and to report on their attitudes and perceptions of the content. For example, such a longitudinal design could be used to test assumptions about reinforcing spirals of political polarization—such that prior attitudes would lead participants to select attitude-consistent content, and exposure to attitude-consistent content would reinforce their attitudes. After inviting panelists to participate in several waves of data collection on selective exposure and reinforcement of attitudes on a given topic, researchers would be able to examine possible reinforcing spiral effects on attitude polarization as assumed in research on the role of echo chambers in political polarization (see Garrett, 2009a, 2009b).

Challenge (4), the study of diverse media channels, is clearly beyond the scope of our present methodological endeavor, the focus of which is on selective exposure to online media. Nevertheless, the possible integration of Piwik with preexisting websites constitutes an important step toward greater external validity of selective exposure research. Piwik enables researchers to extend the scope of possible study designs from the restricted content of mock websites to the study of more diverse content, including regular websites—which allows implementing the selective exposure paradigm in externally valid, naturalistic settings. Note that naturalistic settings usually involve more complex patterns of selective exposure, which often require larger samples to ensure that a sufficient number of participants has, for example, read a specific article. However, the method is also applicable with more strictly controlled designs using mock websites with preselected and experimentally manipulated content (for a discussion of advantages associated with this procedure, see Hastall & Knobloch-Westerwick, 2013). From an ethical perspective, it is important to follow the Web Analyst's Code of Ethics including the five most relevant points that should be clearly communicated to study participants: (1) Participants' personally identifiable data will be safe, secure, and private; (2) data collection practices should be clearly disclosed to study participants; (3) participants should be explicitly empowered to opt out of data collection; (4) participants should be educated about the types of data collected; and (5) the accountability for collected data should be stated clearly to the participants. As Piwik offers monitoring of a wide array of individual web browsing information (such as the IP address), we recommend to set up Piwik in a way that IP addresses are automatically masked or not stored at all. Our ExposureResearchTools purposefully do not export individually identifying, sensitive data beyond the page view activity. Nevertheless, to mitigate privacy concerns of the study participants, it must be clearly stated (i.e., opt-out statement) that study participation includes the consent that browsing behavior will be monitored.

## Limitations

Of course, the present approach to conducting research on selective exposure has several limitations that need to be taken into consideration. Many websites today make use of dynamic content techniques, e.g., scroll content visible when touched by the mouse cursor or display advertising within a layer above the page content. In the default configuration, Piwik will not register any interaction with dynamic content. Also, some web pages contain very much content, allowing users to scroll to the item they are interested in. Again, Piwik will only record that the (large) page was loaded, not which content item was actually displayed on the screen. Such information may be obtained by modifying the JavaScript code provided by Piwik to tell Piwik when specific content becomes visible. It may also be a research interest which button or hyperlink was clicked to reach a specific page of the website. Again, modifications to the JavaScript code (or a workaround with multiple copies of the same page) are necessary to record this information. Such modifications are very specific and go beyond the custom use of Piwik we have presented in this article.

Study participants may not complete the task in a controlled laboratory environment, but on their private devices at home. For those participants who configured their browser to avoid tracking or disable JavaScript, tracking may fail. While this limitation has little effect on samples from the general public, it may substantially affect specific target populations such as people sensitive to Internet privacy or visually impaired users. Researchers may include HTML code in their stimulus to address those participants who have JavaScript disabled (see Figure 4). Furthermore, Piwik records rely on linear use of a web site, that is, one page is read after another. Actually, Internet users may open multiple tabs in their browser and switch among them. Switching among the tabs is not recorded by Piwik by default, because no new content is loaded/displayed. This may lead to inaccurate interpretations of individuals' selection behavior. Again, window/tab switches may be easily tracked via JavaScript. Yet, the necessary modifications, their uses, and the interpretation of window/tab switches regarding selective exposure go beyond the scope of this paper and should be addresses in further research.

## Conclusions

Despite these limitations, we think that Piwik provides a powerful and easy-to-implement research tool that significantly extends the scope of methodological options for selective exposure research. Specifically, Piwik can be used to unobtrusively monitor all relevant aspects of user behavior on any given website, and it can be integrated with custom online survey software. The plug-in ExposureResearchTools takes care of data extraction and preprocessing and thereby considerably reduces the effort to use Piwik for selective exposure research. We hope that the availability of this

```
<noscript>
  <div style="border: 3px solid red; margin: 2em 0; padding: 0.5em;">
    <p>
      <strong>Warning: JavaScript is disabled
      in your Internet Browser.</strong>
    </p>
    <p>Participating in this study requires JavaScript.</p>
    <p>If necessary, please see
      <a href="http://www.enable-javascript.com/" target="_blank">
      this manual</a> on how to enable JavaScript. Thank you.
    </p>
  </div>
</noscript>
```

**Figure 4.** HTML code to address participants visiting the experiment with disabled JavaScript.

research tool will enable and encourage researchers to move beyond the current state of theoretical and methodological study designs of selective exposure research. Among the multiple possibilities is the modeling of selective exposure not only as the dependent or independent variable but also as moderator or mediator of communication processes (see Knobloch-Westerwick, 2015a). A typical scenario for moderating influences of selective exposure would be mutually reinforcing influences (Slater, 2007, 2015) that can best be studied within longitudinal designs. A good example for the mediating role of selective exposure has been mentioned by Knobloch-Westerwick (2015a) who could not show direct effects of exposure to health media content on subsequent behaviors, but effects from individual-level variables on health behaviors were only observable via selective exposure. Hence, integrating selective media exposure with broader models of media effects not only may sharpen our notion of the effect sizes of selective exposure within different media context but also may elucidate how audiences' selective media use influences other media effects. Our hope is to further advance a recent line of methodological and technical development (see Clay et al., 2013; Hastall & Knobloch-Westerwick, 2013), which will eventually make inclusion of selective exposure measures a methodological standard option for media effects researchers.

## Notes

1. If working in a laboratory where multiple participants use the same device(s), we recommend modifying one setting in the default Piwik configuration: the visit timeout. When multiple pages are requested by the same device, Piwik usually assumes that all requests are part of the same "visit" (i.e., viewed by the same participant). Yet, if the lag between two page requests is longer than the visit timeout, Piwik assumes that a new visit (or experiment) has started. The value is specified in a file of the Piwik installation (config/config.ini.php) as *visit_standard_length* with a default value of 1800 (seconds), i.e., half an hour. Depending on the experimental setting, a value of 300 (5 minutes) may be more appropriate. The value should be chosen short enough to be exceeded by the break between two individuals who use the same device. At the same time, the value must exceed the time that a single participant spends on a single web page of the stimulus (time for reading a long article, for instance). In nonlaboratory settings, the default value should do fine.
2. It is important to note that Piwik monitors only "trackable" user actions on a website. By default, opening a web page is such an action, but closing a page is not. In research practice, this especially becomes relevant for the last page that a participant views before leaving the stimulus website and returning to the questionnaire. Given that closing the window creates no measurable user action, the endpoint of the most recent page viewed and thus the respective reading time would be unknown. To handle it, we recommend to either exactly measure the time between opening and closing to stimulus pop-up windows, or to include JavaScript code that notifies Piwik when a window or tab is closed.

## Acknowledgments

The authors wish to thank Sonja Kretzschmar, Verena Wassink, Jennifer Tank, and the Bundeswehr University Munich for sharing information about their ongoing research project with the Federal Office of Civil Protection and Disaster Assistance reported in this paper.

## References

Arceneaux, K., & Johnson, M. (2013). *Changing minds or changing channels: Partisan news in an age of choice*. Chicago, IL: Chicago University Press.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. doi:10.1126/science.aaa1160

Best, S. J., Chmielewski, B., & Krueger, B. S. (2005). Selective exposure to online foreign news during the conflict with Iraq. *The Harvard International Journal of Press/Politics*, *10*(4), 52–70. doi:10.1177/1081180x05281692

Bruns, A. (2005). *Gatewatching: Collaborative online news production*. New York, NY: Peter Lang.

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. New York, NY: Peter Lang.

Bryant, J., & Davies, J. (2008). Selective exposure. In W. Donsbach (Ed.), *The international encyclopedia of communication*. Malden, MA: Wiley-Blackwell.

Bryant, J., & Zillmann, D. (1984). Using television to alleviate boredom and stress: Selective exposure as a function of induced excitational states. *Journal of Broadcasting*, *28*(1), 1–20. doi:10.1080/08838158409386511

Clay, R., Barber, J. M., & Shook, N. J. (2013). Techniques for measuring selective exposure: A critical review. *Communication Methods and Measures*, *7*(3–4), 147–171. doi:10.1080/19312458.2013.813925

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication*, *22*(2), 147–162. doi:10.1080/10584600590933160

Diab, L. N. (1979). Voluntary exposure to information during and after the war in Lebanon. *The Journal of Social Psychology*, *108*(1), 13–17. doi:10.1080/00224545.1979.9711955

Dilliplane, S. (2011). All the news you want to hear: The impact of partisan news exposure on political participation. *Public Opinion Quarterly*, *75*(2), 287–316. doi:10.1093/poq/nfr006

Dillman Carpentier, F., Knobloch, S., & Zillmann, D. (2003). Rock, rap, and rebellion: Comparisons of traits predicting selective exposure to defiant music. *Personality and Individual Differences*, *35*(7), 1643–1655. doi:10.1016/S0191-8869(02)00387-2

Drew, D., & Weaver, D. (2006). Voter learning in the 2004 presidential election: Did the media matter? *Journalism & Mass Communication Quarterly*, *83*(1), 25–42. doi:10.1177/107769900608300103

Feldman, L. (2013). Learning about politics from the Daily Show: The role of viewer orientation and processing motivations. *Mass Communication and Society*, *16*(4), 586–607. doi:10.1080/15205436.2012.735742

Feldman, L., Stroud, N. J., Bimber, B., & Wojcieszak, M. (2013). Assessing selective exposure in experiments: The implications of different methodological choices. *Communication Methods and Measures*, *7*(3–4), 172–194. doi:10.1080/19312458.2013.813923

Fischer, P., Jonas, E., Frey, D., & Schulz-Hardt, S. (2005). Selective exposure to information: The impact of information limits. *European Journal of Social Psychology*, *35*(4), 469–492. doi:10.1002/ejsp.264

Fischer, P., Kastenmuller, A., Greitemeyer, T., Fischer, J., Frey, D., & Crelley, D. (2011). Threat and selective exposure: The moderating role of threat and decision context on confirmatory information search after decisions. *Journal of Experimental Psychology: General*, *140*(1), 51–62. doi:10.1037/a0021595

Frey, D., & Wicklund, R. A. (1978). A clarification of selective exposure. *Journal of Experimental Social Psychology*, *14* (1), 132–139. doi:10.1016/0022-1031(78)90066-5

Garrett, R. K. (2009a). Echo chambers online? Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, *14*(2), 265–285. doi:10.1111/j.1083-6101.2009.01440.x

Garrett, R. K. (2009b). Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication*, *59*(4), 676–699. doi:10.1111/j.1460-2466.2009.01452.x

Garrett, R. K., Carnahan, D., & Lynch, E. (2013). A turn toward avoidance? Selective exposure to online political information, 2004–2008. *Political Behavior*, *35*(1), 113–134. doi:10.1007/s11109-011-9185-6

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, *126*(4), 1799–1839. doi:10.1093/qje/qjr044

Habermas, J. (1989). *Structural transformation of the public sphere*. Cambridge, MA: MIT Press.

Hart, W., Albarracin, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, *135*(4), 555–588. doi:10.1037/a0015701

Hartmann, T. (Ed). (2009). *Media choice: A theoretical and empirical overview*. New York, NY: Routledge.

Hastall, M. R., & Knobloch-Westerwick, S. (2013). Caught in the act: Measuring selective exposure to experimental online stimuli. *Communication Methods and Measures*, *7*(2), 94–105. doi:10.1080/19312458.2012.761190

Hayes, A. F. (2013). Methodology of selective exposure research: Introduction to the special issue. *Communication Methods and Measures*, *7*(3–4), 145–146. doi:10.1080/19312458.2013.845500

Himelboim, I., Smith, M., & Shneiderman, B. (2013). Tweeting apart: Applying network analysis to detect selective exposure clusters in Twitter. *Communication Methods and Measures*, *7*(3–4), 195–223. doi:10.1080/19312458.2013.813922

Jamieson, K. H., & Cappella, J. N. (2008). *Echo chamber: Rush Limbaugh and the conservative media establishment*. New York, NY: Oxford University Press.

Kang, H., Lee, J. K., You, K. H., & Lee, S. (2013). Does online news reading and sharing shape perceptions of the internet as a place for public deliberations? *Mass Communication and Society*, *16*(4), 533–556. doi:10.1080/15205436.2012.746711

Karg, M., & Thomsen, S. (2011). Einsatz von Piwik bei der Reichweitenanalyse: Ein Vorschlag des Unabhängigen Landeszentrums für Datenschutz Schleswig-Holstein (ULD) [Using Piwik for coverage analysis: Recommendations of the Independent Center for Data Protection and Security of Schleswig-Holstein (ULD)]. *Dud - Datenschutz Und Datensicherheit*, *35*(7), 489–492. doi:10.1007/s11623-011-0120-0

Knobloch-Westerwick, S. (2015a). *Choice and preference in media use: Advances in selective exposure theory and research*. New York, NY: Routledge.

Knobloch-Westerwick, S. (2015b). The selective exposure self- and affect-management (SESAM) model: Applications in the realms of race, politics, and health. *Communication Research*, *42*(7), 959–985. doi:10.1177/0093650214539173

Knobloch-Westerwick, S., Dillman Carpentier, F., Blumhoff, A., & Nickel, N. (2005). Selective exposure effects for positive and negative news: Testing the robustness of the informational utility model. *Journalism & Mass Communication Quarterly*, *82*(1), 181–195. doi:10.1177/107769900508200112

Knobloch-Westerwick, S., & Johnson, B. K. (2014). Selective exposure for better or worse: Its mediating role for online news' impact on political participation. *Journal of Computer-Mediated Communication*, *19*(2), 184–196. doi:10.1111/jcc4.12036

Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2015). Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication*, *20*(2), 171–187. doi:10.1111/jcc4.12105

Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research*, *36*(3), 426–448. doi:10.1177/0093650209333030

Knobloch-Westerwick, S., & Meng, J. (2011). Reinforcement of the political self through selective exposure to political messages. *Journal of Communication*, *61*(2), 349–368. doi:10.1111/j.1460-2466.2011.01543.x

Knobloch-Westerwick, S., & Sarge, M. A. (2015). Impacts of exemplification and efficacy as characteristics of an online weight-loss message on selective exposure and subsequent weight-loss behavior. *Communication Research*, *42*(4), 547–568. doi:10.1177/0093650213478440

Knobloch-Westerwick, S., Sharma, N., Hansen, D. L., & Alter, S. (2005). Impact of popularity indications on readers' selective exposure to online news. *Journal of Broadcasting & Electronic Media*, *49*(3), 296–313. doi:10.1207/s15506878jobem4903_3

Knobloch, S. (2002). "Unterhaltungsslalom" bei der WWW-Nutzung: Ein Feldexperiment ["Zig-zagging" towards entertainment in world wide web use: A field experiment]. *Publizistik*, *47*(3), 309–318. doi:10.1007/s11616-002-0068-zn

Kretzschmar, S., & Waßink, V. (2016, June). Individualisierte Risikokommunikation: Audio-visuelle Darstellung von Naturrisiken und ihre Nutzerakzeptanz für den präventiven baulichen Bevölkerungsschutz und den baulichen Schutz kritischer Infrastrukturen [Individualized risk communication: Audiovisual representation of natural hazards and their effects on user acceptance of civil defense measures for the protection of critical infrastructures]. Presentation at the Federal Office for Civil Protection and Disaster Assistance, Bonn, Germany.

Matthes, J. (2012). Exposure to counterattitudinal news coverage and the timing of voting decisions. *Communication Research*, *39*(2), 147–169. doi:10.1177/0093650211402322

Menchen-Trevino, E., & Karr, C. (2012). Researching real-world web use with Roxy: Collecting observational web data with informed consent. *Journal of Information Technology & Politics*, *9*(3), 254–268. doi:10.1080/19331681.2012.664966

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, *41*(8), 1042–1063. doi:10.1177/0093650212466406

Mutz, D. C. (2002). The consequences of cross-cutting networks for political participation. *American Journal of Political Science*, *46*(4), 838–855. doi:10.2307/3088437

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. New York, NY: Cambridge University Press.

Olson, J. M., & Zanna, M. P. (1979). A new look at selective exposure. *Journal of Experimental Social Psychology*, *15*(1), 1–15. doi:10.1016/0022-1031(79)90014-3

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. London, UK: Viking.

Prior, M. (2005). News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science*, *49*(3), 577–592. doi:10.1111/j.1540-5907.2005.00143.x

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge, UK: University Press.

Sears, D. O., & Freedman, J. L. (1967). Selective exposure to information: A critical review. *Public Opinion Quarterly*, *31*(2), 194–213. doi:10.1086/267513

Slater, M. D. (2007). Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory*, *17*(3), 281–303. doi:10.1111/j.1468-2885.2007.00296.x

Slater, M. D. (2015). Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. *Media Psychology*, *18*(3), 370–395. doi:10.1080/15213269.2014.897236

Stroud, N. J., & Muddiman, A. (2013). Selective exposure, tolerance, and satirical news. *International Journal of Public Opinion Research*, *25*(3), 271–290. doi:10.1093/ijpor/edt013

Sweeney, P. D., & Gruber, K. L. (1984). Selective exposure: Voter information preferences and the Watergate affair. *Journal of Personality and Social Psychology*, *46*(6), 1208–1221. doi:10.1037/0022-3514.46.6.1208

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769. doi:10.1111/j.1540-5907.2006.00214.x

Valentino, N. A., Banks, A. J., Hutchings, V. L., & Davis, A. K. (2009). Selective exposure in the internet age: The interaction between anxiety and information utility. *Political Psychology*, *30*(4), 591–613. doi:10.1111/j.1467-9221.2009.00716.x

Wakshlag, J. J., Reitz, R., & Zillmann, D. (1982). Selective exposure to and acquisition of information from educational television programs as a function of appeal and tempo of background music. *Journal of Educational Psychology*, *74*(5), 666–677. doi:10.1037/0022-0663.74.5.666

Wallsten, K. (2005). *Political blogs and the bloggers who blog them: Is the political blogosphere and echo chamber?* Paper presented at the American Political Science Association Annual Meeting, Washington, D.C.

Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication*, *62*(1), 39–56. doi:10.1111/j.1460-2466.2011.01616.x

Zillmann, D., & Bryant, J. (Eds). (1985). *Selective exposure to communication*. Mahwah, NJ: Lawrence Erlbaum.

Zillmann, D., Hezel, R. T., & Medoff, N. J. (1980). The effect of affective states on selective exposure to televised entertainment fare. *Journal of Applied Social Psychology*, *10*(4), 323–339. doi:10.1111/j.1559-1816.1980.tb00713.x