Computers in Human Behavior 27 (2011) 1771–1782

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Flashlight – Recording information acquisition online

Michael Schulte-Mecklenbeck [a,*], Ryan O. Murphy [a,b], Florian Hutzler [c]

[a] Swiss Federal Institute of Technology, Decision Theory and Behavioral Game Theory, Clausiusstrasse 50, 8092 Zürich, Switzerland
[b] The Center for the Decision Sciences, Columbia Business School, 310 Uris Hall, 3022 Broadway, United States
[c] University of Salzburg, Department of Psychology, Hellbrunnerstrasse 34, 5020 Salzburg, Austria

ARTICLE INFO

Article history:
Available online 12 April 2011

Keywords:
Process tracing
Information acquisition
Online research
Decision making
Open source

ABSTRACT

A flashlight enables a person to see part of the world in the dark. As a person directs a flashlight beam to certain places in the environment, it serves as a manifestation of their attention, interest and focus. In this paper we introduce Flashlight, an open-source (free) web-based software package that can be used to collect continuous and non-obtrusive measures of users' information acquisition behavior. Flashlight offers a cost effective and rapid way to collect data on how long and how often a participant reviews information in different areas of visual stimuli. It provides the functionality of other open source process tracing tools, like MouselabWeb, and adds the capability to present any static visual stimulus. We report the results from three different types of stimuli presented with both the Flashlight tool and a traditional eye-tracker. We found no differences measuring simple outcome data (e.g., choices in gambles or performance on algebraic tasks) between the two methods. However, due to the nature of the more complicated information acquisition, task completion takes longer with Flashlight than with an eye-tracking system. Other differences and commonalities between the two recording methods are reported and discussed. Additionally we provide detailed instructions on the installation and setup of Flashlight, the construction of stimuli, and the analysis of collected data.

1. Introduction

A flashlight enables a person to see part of the world in the dark. Our application 'Flashlight' uses this metaphor and shows a participant only a part of the visual stimulus in a research task, while the rest of the stimulus is hidden beneath a blurred layer. Flashlight uncovers a focused area corresponding to the location of the mouse pointer, all the while recording its position and a time stamp. The participants move the mouse pointer as they wish, uncovering different components of interest to them on the stimuli, in the same way a person would direct a flashlight in the dark.

Flashlight is a process tracing tool, i.e., it makes multiple tacit observations of behavior during a research task. This type of process data collection has been of growing interest in recent years. In a review of scientific journal articles on process tracing tools in decision making[1] research, Schulte-Mecklenbeck, Kühberger, and Ranyard (2011) found a steady increase in related publications using a variety of methods during the last 30 years. In that review the authors identify three categories related to studying decision making processes: (1) the tracing of information acquisition, (2) the tracing of information integration and evaluation and (3) the tracing of physiological, neurological, and other concomitants of cognitive processes. For the current paper the first category, information acquisition, is of primary interest. Information acquisition takes place before an actual choice, judgment or rating is made. In these contexts, the researcher is interested in the overall amount of information acquired during a task, the length of inspection time overall and for specific areas of a stimulus and the sequence of information acquisitions. Several established methods like information boards (Payne, 1976; Todd & Benbasat, 1987), eye-tracking (Buscher, Dumais, & Cutrell, 2010; Cutrell & Guan, 2007; Reeder, Pirolli, & Card, 2001; Russo, 1978), active information search (Huber, Wider, & Huber, 1997) or log file analysis (Tauscher & Greenberg, 1997) have been used to investigate (a) psychological processes in information acquisition, (b) human computer interaction (HCI) with computer programs or websites or (c) usability questions in interface design. Different methods were combined to compensate for one methods weaknesses with another method's strengths, e.g., with mouse tracking and log-file analysis (Edmonds, White, Morris, & Drucker, 2007) or thinking aloud and eye-tracking (Eger, Ball, Stevens, & Dodd, 2007).

1.1. Does the hand know what the eyes see?

In what follows we will introduce our method Flashlight that constitutes a form of mouse tracking of participant's information

* Corresponding author. Tel.: +41 2131234323.
E-mail address: research@schulte-mecklenbeck.com (M. Schulte-Mecklenbeck).

[1] Due to our background and interest in decision making we will mostly focus on examples from this research area. Note however that Flashlight was already used in other areas as well, e.g., Boldt, Schulte-Mecklenbeck, and Murphy (2009) used Flashlight for evaluating websites in a usability study.

0747-5632/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2011.03.004

acquisition behavior. Before we do so, we would like to discuss an important question, when it comes to mouse tracking: do traces, collected from mouse movements, represent a participant's information acquisition behavior in an accurate way?

A reference measure for many methodological comparisons in this area, but also in other areas like HCI, seems to be eye-tracking. The reasoning being: if a method shows similar results to eye-tracking it may answer similar questions, but without the substantial costs and technical complexity associated with eye-tracking. Pointing at a target with the mouse has to be slower than pure eye movements towards a target because it always involves both (eye- and hand-movements). There is however evidence for a "strong relationship between gaze position and cursor position" (Chen, Anderson, & Sohn, 2001, p. 281) when web pages are evaluated.

Looking into cognitive psychology we find a similar message that motor movements are not the end product of a cognitive process but are updated "online" (Freeman, Ambady, Rule, & Johnson, 2008; Song & Nakayama, 2006, 2008; Spivey, Grosjean, & Knoblich, 2005).

Given the insights from this literature we want to explore how Flashlight will perform in a comparison with eye-tracking in terms of how accurate it can described how information is acquired from each participant.

1.2. Motivation

The development of Flashlight was motivated by the following aims: (1) we want flexibility in the type of stimuli presented to participants (e.g., overcome the very common matrix setup of stimuli in decision making research), (2) we aim for a cost-effective method which should at best be free of charge, (3) we want to record multiple participants in parallel to minimize experimental trial time, (4) we want data collection within the laboratory as well as online, and (5) we want to enable other researchers to use and modify our code (open-source availability).

In addition to the more programming and software driven aims above we also want to benchmark Flashlight against another process tracing tool in an experiment.

Obviously Flashlight cannot replace eye-tracking, but in several instances it can work sufficiently well (i.e., with a high enough resolution) for researchers who are interested in higher level psychological processes. Moreover it can be implemented without the substantial costs associated with purchasing and operating eye-tracking equipment.

Next we will review several process tracing tools for the laboratory as well as for online research.

1.3. Methods to record process data

We split our review into a laboratory section and an online section. In the laboratory section we focus on eye-tracking because of the close connection to our own tool Flashlight and the large interest that researchers, from a variety of disciplines, have in this method (see e.g., Gompel, Fischer, Murray & Hill, 2007; for a more detailed review of different process tracing tools in decision making see Schulte-Mecklenbeck et al., 2011). In the Web based sections we will introduce mainly open source tools that provide source code without charges but also mention commercial tools to record processes (mostly mouse movements) online. Finally we will compare the described tools on different dimensions.

1.3.1. Laboratory based tools

1.3.1.1. Eye-tracking. The observation of eye movements in psychological research has a long history. An excellent overview over the historical and technical development as well as the basic characteristics of eye movements is provided by Rayner (1998). In contrast to the rather laborious recording of eye movements in earlier days (e.g., Javal, 1878), the technical advances in recent years have resulted in ready-to-use, computer-based eye-trackers. While there are a variety of different methods to measure eye-movements,[2] recording corneal reflection (video based eye-trackers) is the most common method used today (Duchowski, 2002). These high performance eye-trackers record observations often over 1000 times a second and thus deliver data at resolutions of over 1000 Hz. This resolution allows for the measure of both rapid micro-movements (i.e., saccades), as well as fixations (i.e., resting of the gaze on a single location).

1.3.1.2. MouseTracker. MouseTracker (Freeman & Ambady, 2010) records mouse movements produced by participants while they are confronted with visual or auditory stimuli. The purpose of the program is to track trajectories of mouse movements while categorizing stimuli into different classes. Discrepancies between an initial categorization and the final response are shown in deviations from a linear movement of the mouse to one of the response alternatives. MouseTracker comes with a straightforward setup tool, a data recording program, and a package for analyzing and exporting collected data. The validity of the approach has been shown in several publications mainly in the area of stereotyping (e.g. the effects of gender or race; Freeman, Pauker, Apfelbaum, & Ambady, 2010).

1.3.2. Tools to measure processes online

The collection of data via the Internet is well into its second decade by 2010. Multiple websites offer services for online data collection or list ready-made experiments for interested participants. Examples are the 'Psychological Research on the Net' site (Krantz, 2008) which lists over 400 experiments or the 'Web experiment list' (Reips, 2008) with over 500 available studies (Reips & Lengler, 2005). Even large organizations, such as the American Psychological Association (APA), acknowledge online data collection methods in their 'Online Psychology Lab' (American Psychological Association, 2009) a site dedicated to teaching online research methods. Next we briefly review a selection of online process tracing tools that are currently available.

MouselabWeb (Willemsen & Johnson, 2011) is an extension of the well-established Mouselab program for usage on the Internet. Mouselab itself is based on the idea of information boards in which information is presented to the participant in a matrix setup of covered information cells. The mouse is used as a pointing device that automatically uncovers and reveals information in cells upon moving the mouse pointer into a cell area. As soon as the mouse pointer is moved out of the cell area, the information is covered again. Through this process the number of opened cells, their inspection length and sequence can be recorded. MouselabWeb uses Web technology (Javascript, PHP and MySQL) to record information acquisition on the participant's computer. The process data is sent and stored on a server which enables easy, centralized analysis.

WebDiP (Schulte-Mecklenbeck & Neun, 2005) is an online tool that enables a researcher to track participants while they search for information (keywords) in a database. While restructuring of information into matrices could influence the search process, WebDiP lets the participant explore an information space with minimal prerequisites. The system automatically records the order of keywords entered and the information that results from these

[2] e.g., Purkinje image tracking, surface electrodes, search coils (see Rayner (1998), for an overview).

Table 1
Comparison of different process tracing tools.

Tool	Availability	Location	Concurrent participants	Data volume	Flexibility	Setup
Flashlight	Free	Lab/Web	Many	Medium	High	Moderate
MouselabWeb	Free	Lab/Web	Many	Medium	Medium	Moderate
Mousetracker	Free	Lab	Many	Medium	Medium	Moderate
WebDiP	Free	Lab/Web	Many	Low	Medium	Moderate
Eye-tracking	Commercial[a]	Lab	One	High	High	Hard
Medialyzer	Commercial	Web	Many	Medium	Low	Easy
m-pathy	Commercial	Web	Many	Medium	Low	Easy

[a] There are open source eye-tracking systems available, however, most eye-trackers come from a commercial source.

searches. The time spent for task completion can be analyzed as an additional parameter.

On the commercial side there are two noteworthy tools that deliver data closely related to Flashlight: *m-pathy* (http://www.m-pathy.com) and *mediaanalyzer* (http://www.mediaanalyzer.com). Both are positioned in the area of usability testing/market research and collect mouse movements and mouse fixations through different methods. M-pathy uses Javascript code that is embedded in the Website that is tracked. The gathered data are sent back to the companies' website and are analyzed there. Mediaanalyzer collects data not through mouse movements, but mouse clicks. The participant is required to align looking and clicking behavior as closely as possible for the duration of the study.

1.3.3. Comparison of process tracing tools

In this section we want to compare the above introduced tools based on several criteria. This comparison is not a formalized study but a subjective classification that should help the interested reader to be able to have a better differentiation between the available tools.

We selected six criteria for this classification (see Table 1) which we describe in more detail now. *Availability* denotes whether the respective tool is available as open source (full access to the program without attached costs) or can only be bought from a company with restricted access to certain or all parts of the product. Four out of the seven tools are freely available, this group also includes Flashlight. Eye-tracking systems mostly come from a commercial vendor, therefore we classified them together with Medialyzer and m-pathy, note however that there are projects available that aim at building free eye-tracking system (e.g., Babcock, Li, Parkhurst, & Winfield, 2010). The second criterion denotes the *location* where a tool can be used. Mousetracker is laboratory bound through the construction of the software and eye-tracking systems also fall in this category because of the hardware need to record data. Medialyzer and m-pathy can only be used through the Internet and the respective companies' web site. For the remaining tools both locations are possible through installing Flashlight, MouselabWeb or WebDiP on a web server and providing access for participants in a laboratory.

When testing participants an important economical question is how many *concurrent participants* can be run in a session. This criterion clearly differentiates eye-tracking from all the other reviewed methods. Only one participant per session can be run with an eye-tracker in a lab setup. For the other lab bound methods this number is determined by the number of computers available in a lab or without a limitation when it comes to Web bound methods.

The category *data volume* investigates how much information per participant and time frame each methods delivers: eye-tracking provides the largest amount of data and highest time resolution with at least several megabytes of data per participant and often over 1000 Hz sampling resolution. WebDiP stores the least amount of information with a few kilobytes per participant and a time resolution on the level of seconds.

Flexibility describes how flexible a researcher is in terms of stimuli used in the different methods: Flashlight and eye-tracking systems can use any visual stimulus and therefore lead this category, the commercial systems Medialyzer and m-pathy depend on Webpages as stimuli and are therefore less easy to manipulate. Mousetracker has the advantage of being able to present visual as well as auditory stimuli but focuses on simple images for categorization. MouselabWeb and WebDiP come with relative restricted options to depart from a standard setup (information boards or a search engine interface, respectively).

The final category, *setup*, describes the simplicity and speed of setting up a study: as most eye-trackers demand a calibration process and some apparatus to be setup in a laboratory they get a low score here, Flashlight, MouselabWeb, WebDip and Mousetracker require a Webserver to run and each come with their own configuration tools (either in the form of configuration text files or additional software to design experiments). Setting up a Webserver demands some technical knowledge, therefore we categorize these tools in the 'moderate' group. Medialyzer and m-pathy can be setup through the companies respective websites and require no additional tasks, a process we categorize as 'easy.'

To summarize, eye-tracking systems deliver detailed data and flexibility in terms of what is shown to participants as stimuli. The systems are relatively cost intensive and data collection can only be done one participant at a time. In many laboratories they are the standard tool to collect process data. The other available tools are all relatively new and hence have to build up a larger user base. Time will show which of them will play an important role in the recording of information acquisition processes.

Flashlight comes with the same advantages of flexibility in terms of stimuli as eye-tracking, yet the program has no associated acquisition costs and can run many participants at the same time in a lab or via the Internet. The downside is a lower sampling frequency and hence restrictions in terms of observations made in eye-trackers, e.g., saccades, that are not available in Flashlight. In this sense, Flashlight is not a replacement for eye-tracking measurement in total, but can be used in lieu of eye-tracking for the subset of research projects where macro level behaviors are of primary interest.

1.3.4. Limitations of the currently available online tools

As outlined above, we briefly discussed several tools available to collect process data online (Flashlight, MouselabWeb, WebDiP, m-pathy and mediaanalyzer). In what follows we want to extend the comparison above with a list of limitations of each methods and link those limitations to remedies offered by Flashlight.

The information structure in a Mouselab(Web) setup is in a grid form. This presentation restriction may influence how participants acquire and use information (see Lohse & Johnson, 1996), for a detailed discussion on this issue). We know that most Westerners have a left-to-right, top to bottom manner in acquiring information and the use of a grid structure for organizing information may exacerbate this.[3] Although simple counterbalancing of the matrix orientation can lower the influence of these standard acquisition patterns it cannot be ruled out that there remains a left–right dominance in the acquisition data. The labeling of the columns and rows, as well as the cells of such information matrices provides additional information to the participant that is not available in natural decision environments (see, e.g., the Active Information Search paradigm for a different approach, Huber et al., 1997). Acquisition of labeling information cannot be observed in MouselabWeb[4] because only cell openings as such are recorded. On a process data level only binary information about a cell opening or closing is available. MouselabWeb is therefore blind to finer grained movements within and between cells that may be of interest to researchers. Further, MouselabWeb cannot accommodate non-matrix information structures (e.g. a webpage layout, a food product label, an organizational chart) that may be of interest to researchers as stimuli.

WebDiP overcomes the problem of pre-structuring. During the search process the participant enters a keyword into the search window and gets a list with questions related to this keyword (this resembles the results section of a Google or Bing search to some extent). Then the participant selects one of the questions by clicking on it and receives an answer. While all this is traceable, the reading process of the results section stays oblique. It remains unclear whether the participant reads only the first two or three results or goes through the whole list before clicking on a question.

The basic principle of m-pathy is closely related to Flashlight, the downside of this tool is the obligation to use the company's webserver to collect data. This might cause problems with, e.g., confidentiality or possible unavailability of the company due to economic reasons. Additionally in a closed source tool like m-pathy the steps of collecting data and analyzing them remains oblique to the researcher.

Finally, regarding mediaanalyzer, it remains unclear how constant clicking to an area, the participant is attending to in mediaanalyzer, changes the actual acquisition process. The mediaanalyzer company cites an internal study where high validity of the method is assured, however (as with m-pathy) no public control of this assertion is provided.

Flashlight offers several remedies to the just described restrictions. First and foremost the information structure in Flashlight is completely flexible and allows for any visual representations of information. Any static stimulus material (text, tables, pictures, screen shots, cards, etc.) can be displayed and used; it is important to note that dynamic stimuli cannot be applied in the current version. The Flashlight representation of information is therefore more flexible and natural than a MouselabWeb setup. The process of gazing and bringing into focus areas of interest is natural in human information acquisition and is captured well by mouse movements as demonstrated by multiple studies (e.g., Chen et al., 2001; Spivey & Dale, 2006). The Flashlight system is able to record up to 10 observations a second of where participants are moving their mouse. From these data researchers can infer what information participants acquire, the velocity of movements between information bits, as well as subtle movements between areas of interest. Finally, the main difference to the commercially available tools, m-pathy and mediaanalyzer, is that Flashlight is fully open-source which means that anybody can download and modify the whole

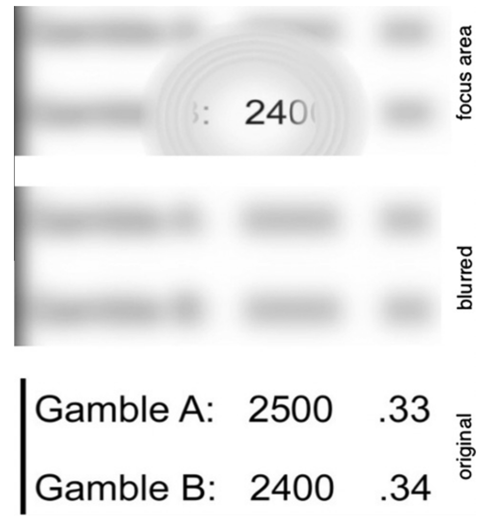

Fig. 1. Construction of a stimulus with the original layer at the bottom, the blurred version in the middle and the final version with the focus area at the top.

program code for free and check the analysis and recording scripts within the project.

In what follows this we will introduce technical aspects regarding Flashlight in more detail. Then we will show results from an experiment using Flashlight in one group and eye-tracker in the other with three different stimuli types. We conclude with a discussion of this new tool's potential uses. Several detailed appendixes are included providing technical details about installing and configuring Flashlight.

2. Flashlight – technical aspects

Stimuli in Flashlight always consist of two parts: a bottom layer constituted by the original stimulus and a top layer, a blurred version of the stimulus. In Fig. 1 we present an example of the original stimulus of a gamble at the bottom, the blurred version of this stimulus in the middle and the blurred version + the focus area at the top of the figure.

Upon starting a trial the participant sees only the blurred version (middle part of Fig. 1) and the mouse cursor with a surrounding circular area[5] which we call the focus area (see Supplementary material: Configure Flashlight for more details on how to configure different types of focus areas). Like a flashlight beam controlled by the participant, moving the mouse over the stimuli reveals clearly and instantly the underlying bottom layer within the focus area (top part of Fig. 1). Code written in Javascript takes care of this revelation process, revealing information where the mouse is pointed and obfuscating the remaining field of the stimulus, all the while updating in real time as the participant moves the mouse. Additionally this program records the position of the focus area at a rate of about 10 times a second (10 Hz resolution) and saves the collected data together with a timestamp into a database for analysis. The temporal resolution can be increased beyond this level but is ultimately limited by the speed of the client computer and the capabilities of Javascript to record mouse positions. In any case, for a large number of research instances a 10 Hz resolution is sufficient to measure participant's behaviors, especially in judgment and choice tasks.

The whole source code plus scripts for analysis are available at the projects webpage: http://vlab.ethz.ch/flashlight. We also

[3] Glöckner and Betsch (2008) generated a presentation format in which the standard matrix setup of rows and columns is replaced with circular setup that has a central starting point and equal distances to all presented information.

[4] One way to get insight into the usage of column or row labels with MouselabWeb would be to cover those labels, too. While this is possible, the search experience might be changed by such an approach.

[5] Clearly Flashlight deprives the user of peripheral visual information always available in an eye-tracking environment available (without additional cost for the participant). We will come back to this point in the discussion section.

provide additional instructions for installation (Supplementary material), stimulus generation (Supplementary material) and the analysis of the collected data (Supplementary material) at the end of this paper.

3. Experiment

To validate the data collected in Flashlight we conducted an experiment that compares results from the same stimulus set presented in Flashlight versus a conventional eye-tracker. The decision to use eye-tracking as a comparison was based on the following reasons: (a) in the literature eye-tracking plays an important role in the evaluation of process tracing tools (e.g., Abelson & Levi, 1985; Lohse & Johnson, 1996), (b) eye-tracking is often seen as the "gold standard" in process tracing because of its unobtrusive means to collect information acquisition with high accuracy and freedom in stimulus design, (c) in a literature review and analysis in Kühberger, Schulte-Mecklenbeck, and Ranyard (2011), eye-tracking methods showed a steady increase in usage over the last 10 years (likely due to the fact that eye-tracking setups became more sophisticated and streamlined, hence easier to use, over time).

As already discussed in the comparison of different process tracing methods, we are aware that there are differences between the methods that might influence the results to a certain degree. For example the time needed to move from one information item to the next in Flashlight is necessarily larger than the time a simple eye-movement between the same information items takes. Saccades, which can easily be picked up by eye-trackers are below the registration threshold of Flashlight. However, as these two methods constitute the cornerstones of our above review, we aim to describe the differences and evaluate where each methods has its strengths and weaknesses. Empirically, as both eye-tracking and Flashlight can measure attention and information acquisition, we expect generally not to find significant differences between the two methods. Moreover we do not expect people to make different decisions based on the measurement method.

3.1. Method

We compared the two methods (Flashlight (FL), eye-tracking (ET)) in a between subjects design using three different tasks as independent variables. These tasks were an arithmetic task (adding five numbers, i.e.: add task), a risky choice task (deciding between two gamble options, i.e.: choice task) and a reading task (reading a novel paragraph of text, i.e.: read task). The same stimuli were used in three tasks in both samples; in FL as the bottom layer (see Supplementary material for details on the stimulus construction in FL) with a blurred version as the top layer and in ET displayed "as-is." The three tasks (add, choice and read) differed in their complexity (in terms of AOIs), their mental effort and their theoretical base. The add task represents a simple stimulus structure (see Supplementary material for an example in Fig. C1) with five AOIs (Areas of Interest) indicated by the five numbers. The participant had to add up the five numbers and indicate the result. In this task we are able to track acquiring numbers in a small amount of AOIs and adding them up, which adds an accuracy measure. The choice task contains two two-options gambles represented in a more complex stimulus structure with eight AOIs (see Supplementary material for an example in Fig. C2) indicated by the eight numbers in the display (outcome or probability). This task is based on gambles from Prospect Theory (Kahneman & Tversky, 1979) and follows the idea that people weight the outcomes of each option by the adjacent probability, i.e., they first multiply the two values and compare them in a second step, hence acquisition patterns that concentrates on within option transitions should be found. The Priority Heuristic (Brandstätter, Gigerenzer, & Hertwig, 2006) makes the opposite prediction in comparing outcomes and probabilities between options separately, hence between options transitions should be found (see Johnson, Schulte-Mecklenbeck, & Willemsen, 2008 for a more detailed discussion of the different predictions of the two theories). In the read task we selected hard to comprehend texts (see Supplementary material for an example in Fig. C3) in order to see how FL performs in comparison to ET in a very demanding task (in terms of process tracing).

3.1.1. Apparatus

Flashlight data were recorded with the freely available Flashlight code, Version 0.8. Eye movements were recorded using a video-based remote iView X RED-III system (SensoMotoric Instruments) with a temporal resolution of 60 Hz. Stimuli were presented on a 20-in. CRT monitor connected to a Windows2000 compatible computer. Stimulus presentation was controlled by Presentation (Neurobehavioral Systems) software.

3.1.2. Procedure

In the FL sample participants were introduced to the basic handling of the FL system and the usage of the computer mouse on separate HTML pages. Each trial after this introduction started right away without any further delay or calibration procedures (see below). In the ET condition the eye-tracking system was first calibrated individually for each participant. After successful calibration, requiring between 2 and 3 min, each trial started with the presentation of a fixation cross in the middle of the screen. The cross remained on the screen until it was fixated upon by the participant (or up to 5000 ms in case of poor eye-tracking calibration, resulting in an automatic calibration verification and, if necessary, a re-calibration, again for 2–3 min). Subsequently, the experimental stimulus was presented and remained on the screen until a button was pressed by the participant. Each experimental stimulus was followed by a brief procedure (two fixation crosses had to be fixated sequentially, one in the left and one in the right half of the screen) verifying the proper calibration of the eye-tracking equipment during the experiment. Given that during this procedure poor calibration was recognized, a re-calibration was automatically initiated. While especially this last step is not essential when conducting an ET study, it ensures that data quality is stable during the whole recording session.

In both samples (FL, ET) participants did an independent warm-up task to familiarize them to the respective system. This step is especially important in FL due to the little information the participant can initially see on the blurred stimulus. Getting accustomed to the usage of the Mouse and the focus area are additional steps to ensure a smooth data collection during the actual trials used in the analysis later. Following the warm-up task seven stimuli from each type (add, choice and read), were presented within blocks in random order; the blocks themselves were also presented in a random order to each participant. In the 'add' and 'choice' stimuli the different information items (numbers in the add task, outcomes and probabilities in the choice task) were additionally counterbalanced in terms of presentation order (top, middle or bottom row in the add task; top or bottom row, left or right column in the choice task) to avoid reading order effects.

After each 'add' stimulus the participants were asked to enter the sum of the presented numbers, after each 'choice' stimulus they were asked to specify the gamble they would like to play and after the 'read' stimulus they were asked to answer a general knowledge question about the just read text.

3.2. Participants

The FL sample, drawn form the Bergen virtual laboratory panel (http://vlab.uib.no, which recruits locally at the university but is also open to outside participants), consisted of 23 female (mean age of 37.9 year) and 27 male (mean age 29.1 years) participants. The ET sample was drawn from the local student community at the University of Bergen and consisted of 13 female (mean age of 25.4 year) and 9 male (mean age of 24.3 year) participants (from two additional participants no data were collected due to problems getting a good calibration in the eye-tracking system). Participants were compensated with a fixed payment of US $6.00 each for their time and efforts.

3.3. Dependent variables

We will focus our analysis on two data types: *outcome results* and *process results*. Outcome results are task specific; they include the actual entered numeric result of the addition in the add task, the chosen option in the choice task and the answer to a question about the text in the reading task. For the add (correctness of addition results) and the read task (correctness of response to question regarding the text), the outcome results will provide a measure of accuracy of the participants' responses and hence detect possible method driven outcome differences between ET and FL. The outcome results of the choice task are additionally informative because opposite acquisition patterns should be found based on prospect theory (Kahneman & Tversky, 1979, more within option comparisons) and the Priority Heuristic (between options comparison, Brandstätter et al., 2006).

For the *process results* several dependent measures are useful: The *time* participants need for completing a task is a general measure of task complexity and processing depth. Task completion time should increase with task complexity as well as with the amount of information in a task. The next level, *fixations*, informs us about the distribution of fixations over the different parts of the stimuli. Consistent with standard practice in eye-tracking research, we defined a fixation as a period in which the retina is stabilized over a stationary object (Duchowski, 2007). We use the same logic for FL: if the focus area does not move more than 25 pixels within a given time frame, we would identify this as a fixation event. The center of the fixation event is defined as the point that is the geometric center of all the focal points produced by a participant during the fixation. This measure will also correlate positively with the amount of information in a stimulus but moreover show where the information acquisition of the participants is directed. On a more detailed level, *fixations per AOI* isolate the observations of different facets of information (defined by the researcher) contained in the stimuli. The *pattern* and *sequence* of acquired information will serve as a fourth dependent measure in the process results. The sequence of fixations informs us about how different pieces of information are used during the information search process. Measuring acquisitions of information in, e.g., the choice task allows for the testing of detailed predictions of decision processes. This dependent measure also lends itself for visualization by using methods like heat maps and transition graphs (see results section for examples).

3.4. Hypotheses

We formulate two sets of hypotheses, one based on outcome measures and one based on process measures.

3.4.1. Outcome hypothesis

Add/Reading task hypothesis between methods: For the add and reading tasks, neither task nor method differences should arise.

Choice task hypothesis within methods: In the choice task we predict that the subset of gambles with the same expected value (gamble 1–4, see Supplementary material, Table C1) will result in indifferent choices close to 50% whereas the second subset with different expected value (gamble 5–7) the choices should favor the option with the higher expected value.

Choice task hypothesis between methods: The above described choice patterns should be found regardless of the used method (FL and ET).

3.4.2. Process hypotheses

The first set of hypotheses covers task length within and between the two examined methods FL and ET.

Task length hypothesis within methods: We expect an increase in task completion time across our three task types (add, choice and read). Add should be completed fastest, choice should be situated between add and read, read should result in the longest completion times. The predicted pattern is based on the amount of information to be read and the operations necessary to come to an answer or decision. For example in the add task 5 numbers have to read and added up, whereas in the choice task 8 numbers have to read and a much more complicated evaluation has to be performed (assuming a 'normative' solution requires multiplying each outcome and probability pair, adding these two numbers for each option and comparing the results. Note that assuming a more heuristic search pattern which includes less operations to come up with a choice, necessarily includes the reading of the eight numbers as a minimum acquisition pattern).

Task length hypothesis between methods: Across the three tasks the ET condition should result in shorter completion times than the FL condition. This is due to the additional motor movement resulting from the use of the mouse in FL.

The next set of hypotheses covers length of fixations and number of fixations as an indicator for information acquisition behavior.

Fixation hypothesis within methods: We expect an increase in the number of fixations across our three task types (add, choice and read). This prediction is based on the increase in the number of AOIs from add (5) to choice (8) and read (whereby during 'read' AOIs are equivalent to what is commonly referred to as 'number of fixations' in the literature on eye movements in reading).

Fixation length between methods: Transition hypothesis for choice: For the choice task we predict more within option transitions than between option transitions following a compensatory pattern of information search (this reasoning goes back to results found in Johnson et al. (2008)). Generally we expect the two methods to uncover the same basic psychological processes. We expect parallel results in the amount of fixations exhibited and similar patterns of where participants' acquire information.

4. Results

In what follows we will first inspect the results for the outcome measures for the three tasks add, choice and read. Then the following process measures will be reported: the overall time for task completion, the number of fixations, the duration of fixations, the pattern of fixations, and lastly the structure of the transitions between fixations within areas of interest.

4.1. Outcomes measures for the add, read and choice tasks

For each task type different behavioral measures were calculated: (1) the accuracy in the adding task (percentage of correct answers in the addition task), (2) the choice pattern in gambles with the same versus gambles with different expected values, (3) the accuracy in the reading task (percentage of correct answers).

For add we find a higher percentage correct in the ET (75.97%) condition than with FL (68.42%, $z < 1.96$, n.s.). In reading this pattern is reversed with more correct responses in FL (71.71%) than

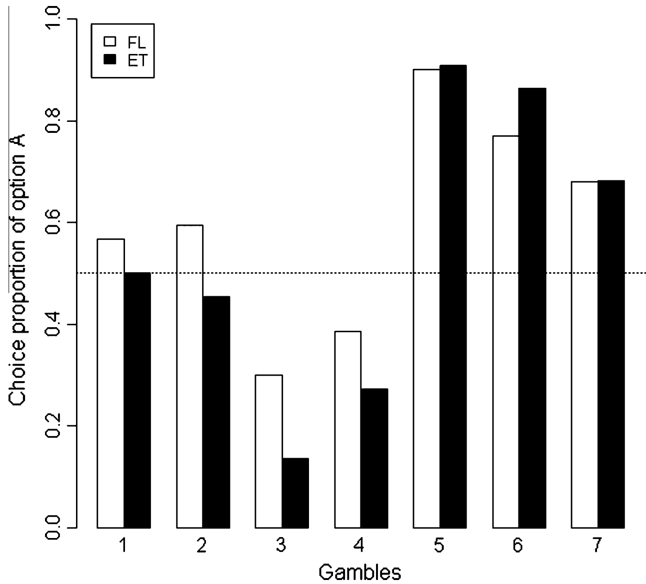

Fig. 2. Choice patterns for the choice task. Gambles 1–4 with equal expected value, gambles 5–7 with higher expected value for option A.

in the ET (64.23%, $z < 1.96$, n.s.) condition. However, none of the proportions are significantly different at the .05 level when comparing across the FL and ET conditions using two sample proportion tests.

For gambles with equal expected value (gambles: 1–4 in Supplementary material, Table C1) participants should be indifferent between the gamble options. For gambles with different expected values (5–7 in Supplementary material, Table C1) the normative choice rule we explore is, choose the gamble with the higher expected value, in all three gambles the option A. Gambles 1 and 2 result in a 50:50 distribution (see Fig. 2) of the choices (implying indifference between gamble A and B) and no significant difference between the FL and the ET sample using two sample proportion tests ($z_{G1} = 0.195$, $p < .85$; $z_{G2} = 0.731$, $p < .46$, two-tailed). Results from gambles 3 and 4 show similar patterns between the two methods no significant differences are found between the ET and FL sample ($z_{G3} = 0.966$, $p < .33$; $z_{G4} = 0.512$, $p < .61$, two-tailed). For the second group of gambles, with different expected values, a clear preference for option A can be found with only slight variations between the two methods. In these gambles (5–7) participant's choices generally follow the normative solution of choosing the gamble with the higher expected value. No significant differences were found between methods (all $z < 1.96$, n.s.).

4.2. Process measures for the add, read and choice tasks

4.2.1. Fixations

The average number of fixations per task are presented in Fig. 3 (left half).

The number of fixations differ between the three tasks with the smallest amount in the add task (average across methods: 26.53), an increase in the choice task (average across tasks: 36.35) and the maximum number of fixations in the read task (average across tasks: 111.53). As hypothesized, ET and FL methods resulted in comparable numbers of fixations on each task, except for significantly more fixations ($p < .01$) on the choice task. This is the only significant difference that was detected with a set of non-parametric Kolmogorov–Smirnov[6] tests controlling for familywise error. No other significant differences were found with this dependent variable. This increase in number of fixations is consistent with our hypothesis that the number of fixations should increase as stimulus complexity increases.

4.2.2. Durations of tasks

The read task took the longest to complete with an average of 43.27 s across methods. While we expected this result, it was to our surprise that the choice task had the fastest mean completion time (average across methods: 12.02 s) and the add task took participants nearly the double length (average across methods: 22.98 s). See Fig. 3 (right half) for task duration for each method by task combination. Regarding our hypothesis this indicates that we did not find an increase of task length in the strict pattern of add < choice < read, but did find a reverse in longer completion times for add than for choice.

4.2.3. Patterns of fixations

The pattern of participants' fixations can be represented effectively as a heat map, a 2-dimensional figure containing gradations in color that correspond to different values at different locations in a 2-dimensional coordinate system. In this case, brighter areas indicate more fixations whereas darker areas are those areas with fewer fixations. Fig. 4 shows the pattern of information acquisition, as manifested by fixation points, across all participants in both the ET and FL condition for the add and choice tasks. Example stimuli are overlaid in the figure to provide a spatial reference. As is clear when examining the figure, the resulting patterns of acquisition are similar between the two data collection methods. One noteworthy difference is in that the FL method picks up on the movement between the different stimuli better than eye-tracking. The reason is simple in that a mouse's movement is continuous whereas eye fixations may be discontinuous (saccadic) and prone to greater accelerations and decelerations than the manual movement of a computer mouse.

4.2.4. Fixations and areas of interest

Particular Areas of Interest (AOI) can be defined that correspond to the underlying structure of the stimuli. For example, the choice task had eight pieces of information arranged in two rows and four columns. The 'cells' of information can be isolated as polygons and fixation points that fall within each of these areas identified (see Fig. 5. Areas of interest (in gray) for the choice task with Flashlight including the collected data fixation as points). These AOIs can serve as a framework for subsequent analyses. For example, one could isolate the number of fixations in each AOI or develop models of the transitions between different AOIs. We will give examples for both approaches now.

First, considering the choice task, we compare the distributions of the frequencies of fixations, organized by areas of interest, across the two methods. Table 2 shows the percentage of fixations in each of the areas of. We labeled the eight AOIs based on the type of information (Outcome value or Probability value), the gamble (a or b) and the options (option 1 indicating the left column and option 2 indicating the right column). This means the first outcome of gamble a is coded as Oa1 and the second probability of gamble b coded Pb2 (we added these labels in Fig. C2 in Supplementary material).

The first option pair of each gamble (Oa1, Pa1 and Ob1, Pb1) receives the largest interest in ET as well as FL. In ET the distribution of interest drops the further participants navigate within one gamble (on average from 17.81% for Oa1 to 9.52% for Pa2; from 13.47 for Ob1 to 5.93 for Pb2, overall for all the AOIs a range of 11.88). This means a range of 8.29% points for gamble a and a range of 7.54% points for gamble b. Inspecting the same gambles in FL we see less focus in the information search patterns across the

[6] Nonparametric significance tests were used because the assumptions of normality were not met.

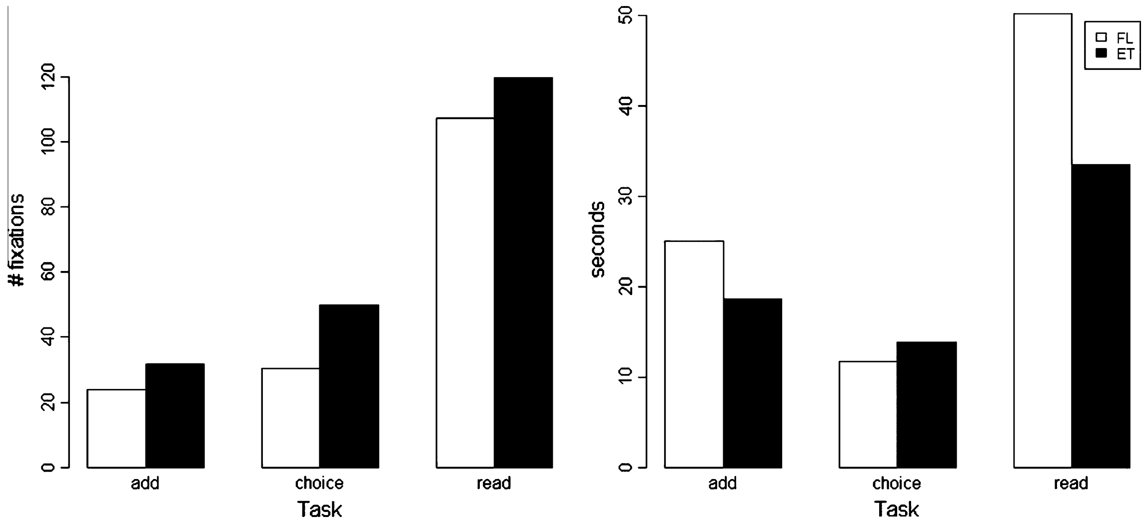

Fig. 3. Mean completion time in seconds (left half), mean number of fixations (right half) for Flashlight (FL) and eye-tracking (ET).

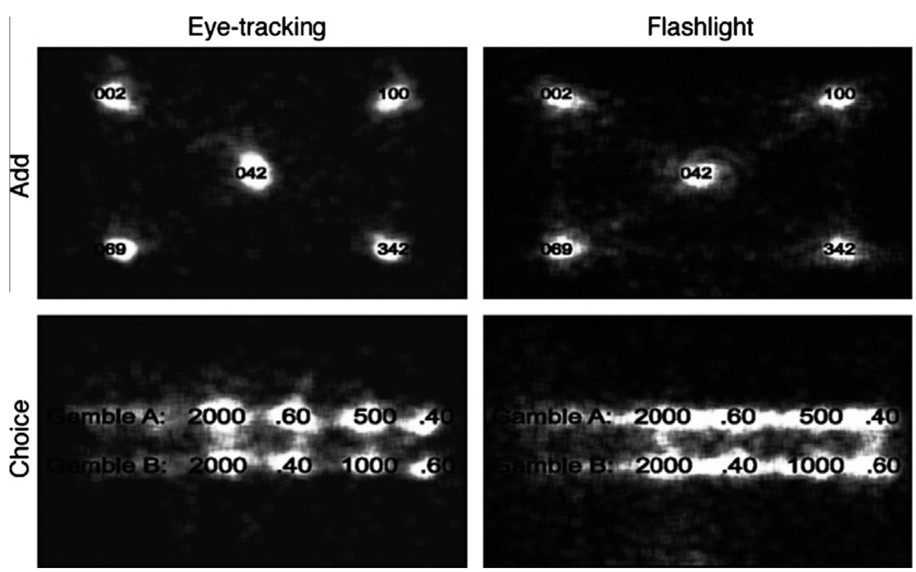

Fig. 4. Heatmap for the add (top row) and choice (bottom row) task for Flashlight (left column) and Eye-tracking (right column) methods.

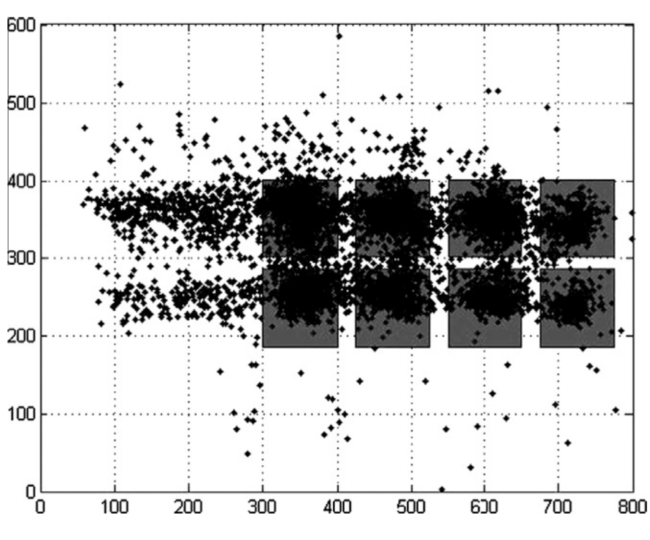

Fig. 5. Areas of interest (in gray) for the choice task with Flashlight including the collected data fixation points.

Table 2
Percent of fixations in the choice task: Eye-tracking and Flashlight in the eight Areas of Interest (AOI).

AOI	Eye-tracking	Flashlight
Oa1	17.81	14.19
Pa1	17.37	12.67
Oa2	14.91	11.04
Pa2	9.52	11.26
Ob1	13.47	14.04
Pb1	11.51	12.35
Ob2	9.44	11.15
Pb2	5.93	13.25

different AOIs with range differences of 2.93% for gamble a and 2.89% for gamble b (overall for all the AOIs a range of 3.15).

A separate analysis included the area around the eight AOIs and indicated that indeed the majority of fixation points fall within one of the AOIs (69.60% in ET and 61.03% in FL). That the FL method has more fixations points (38.97%) outside of one of the eight defined AOIs is not a surprise, as in this condition participants were required to move their mouse pointer to the bottom middle of the screen to register their choice whereas ET participants simply pressed a key on the keyboard indicating their choice. These procedural points can be seen in Fig. 5 with careful inspection. Concluding this part of the AOI analysis we find that none of the differences between AOIs were significant, which supports our claim that FL shows convergent and concurrent validity.

4.2.5. Transitions between AOIs

While predictions are also possible for the other tasks they are much clearer in the choice setup because the order in which people attend to information in risky choices has been the topic of recent interest (Brandstätter et al., 2006; Johnson et al., 2008). We will use the gamble task to demonstrate this point, because we can make initial theoretical predictions that we can then test. Standard expected value models predict a within option transitions pattern integrating each outcome/probability pair (e.g., Oa1 and Pa1). This pattern is indicative for a compensatory search strategy. Models based, e.g., on the priority heuristic (Brandstätter et al., 2006), predict a between option transition pattern indicative of a comparison, e.g., based on the maximum outcomes of both gambles only (e.g., Oa1 and Ob1). We use a simple test to inspect these predictions: After participants view Oa1, what information are they most likely to investigate next? Generally, the order that people shift their acquisitions between different AOIs can be modeled as a hidden Markov model. We follow now by deriving from experimental data two transition matrices (one from FL and the other from ET) of participant's shifts of acquisition between different elements of the stimuli (see Tables 3 and 4). We marked the highest transition probability per row bold in both matrices, e.g., in the first row in Table 3 this value is 0.64 for the transition between Oa1 and Pa1. These transitions matrices show that the highest transition probability is in most cases found between adjacent cells, within gambles, in both directions, AOIs: Oa1 to Pa1; Pa1 to Oa1; Oa2 to Pa2 and so on. In the ET method this observation is true for all adjacent AOIs, in the FL method it holds for all but the transition between Ob1 and Oa1 which is more frequent than the expected Pa1–Oa1 transition. The exception to this pattern is the transition between AOIs: Oa1 to Ob1 and Pa2 to Pb2, these two transition patterns indicate the switch from one gamble to the other and back. The results further show that these transition matrices, although derived from different methods and participants, are similar. The mean absolute difference between the two transition matrices is .044; this indicates that overall the different transition matrices are roughly equivalent to each other—an absolute difference of exactly 0 would indicate perfect correspondence whereas randomly different transition matrices with similar probability distributions would yield a mean absolute difference of about 0.150. The probability of observing a mean absolute difference as low as 0.044 by chance alone is very small ($p < .001$). See Bartlett (1951) for a discussion of similar goodness of fit tests.

Transition matrices can be graphically represented as arrows (indicating direction of transitions) in different thickness (indicating probability of transitions) between AOIs. For an example see Fig. 6 where we depict transitions from Oa1 to the other seven AOIs (top row) and transitions from Pa1 to the other seven AOIs in the bottom row as an example. The strongest transition is found between Oa1 and Pa1 (as already indicated in the transition matrix results above) all other transitions have much smaller probabilities which can be seen in the much thinner arrows. Comparing the two methods ET (left column) and FL (right column) the independence of the transition probabilities from the used method becomes obvious – both pictures resemble the nearly the same transition patterns.

Table 3
Eye-tracking AOI transition matrix for the choice task.

AOI	Oa1	Pa1	Oa2	Pa2	Ob1	Pb1	Ob2	Pb2
Oa1	–	**0.64**	0.15	0.02	0.13	0.04	0.02	0.00
Pa1	**0.38**	–	0.30	0.09	0.07	0.10	0.04	0.01
Oa2	0.11	0.19	–	**0.42**	0.08	0.04	0.13	0.03
Pa2	0.05	0.14	**0.50**	–	0.04	0.09	0.07	0.12
Ob1	0.29	0.04	0.07	0.01	–	**0.48**	0.10	0.00
Pb1	0.04	0.19	0.02	0.03	**0.46**	–	0.20	0.05
Ob2	0.05	0.06	0.20	0.02	0.15	0.12	–	**0.41**
Pb2	0.03	0.06	0.08	0.14	0.06	0.16	**0.46**	–

Table 4
Flashlight AOI transition matrix for the choice task.

AOI	Oa1	Pa1	Oa2	Pa2	Ob1	Pb1	Ob2	Pb2
Oa1	–	**0.60**	0.07	0.02	0.28	0.02	0.01	0.00
Pa1	0.24	–	0.33	0.13	0.04	0.21	0.02	0.03
Oa2	0.06	0.14	–	**0.54**	0.03	0.03	0.15	0.04
Pa2	0.04	0.08	**0.34**	–	0.03	0.02	0.04	0.46
Ob1	**0.26**	0.11	0.02	0.02	–	**0.49**	0.07	0.03
Pb1	0.03	0.20	0.06	0.02	**0.37**	–	0.25	0.06
Ob2	0.03	0.01	0.13	0.05	0.08	0.26	–	**0.45**
Pb2	0.01	0.02	0.06	0.27	0.09	0.11	**0.44**	–

5. General discussion

The idea for Flashlight grew from the observation that there are no open source tools that collect information acquisition data with high flexibility for stimulus selection and the option to test multiple participants concurrently. In developing Flashlight we were inspired by the flexibility of eye-tracking systems and the acquisition process of different mouse tracking tools. In what follows we will first discuss a methodological issue regarding methods' comparison and why we chose eye-tracking as our reference method, we will move onto compare FL and ET in reference to the data we collected and provide cross method evidence for the validity of our approach and finally point to possible other applications outside our own field.

5.1. What to compare to?

In our study we decided to compare Flashlight to eye-tracking. As mentioned in the introduction section, there are many different process tracing tools available – eye-tracking is the most appropriate one for our situation because it offers the same flexibility in the type of stimuli that can be used and is the reference point for comparisons in many publications. The second closest candidate was MouselabWeb, we decided against this tool, because we would have not been able to use our read task.

5.2. Flashlight versus the eye-tracker

We compared data collected online with Flashlight to data collected in the laboratory with a conventional eye-tracking system using the same task set. There is an inherent confounding between method and research location that we cannot resolve in the current research setup. An alternative design would treat the two methods as within-subjects-factors, comparing the same tasks. This would have solved problems connected to demographic differences in the two samples we collected, however introduce an issue of redoing the same task twice and also limiting our sampling to the local student population.

It is also apparent that there are differences between the two methods in the way information is acquired. Eye-tracking profits

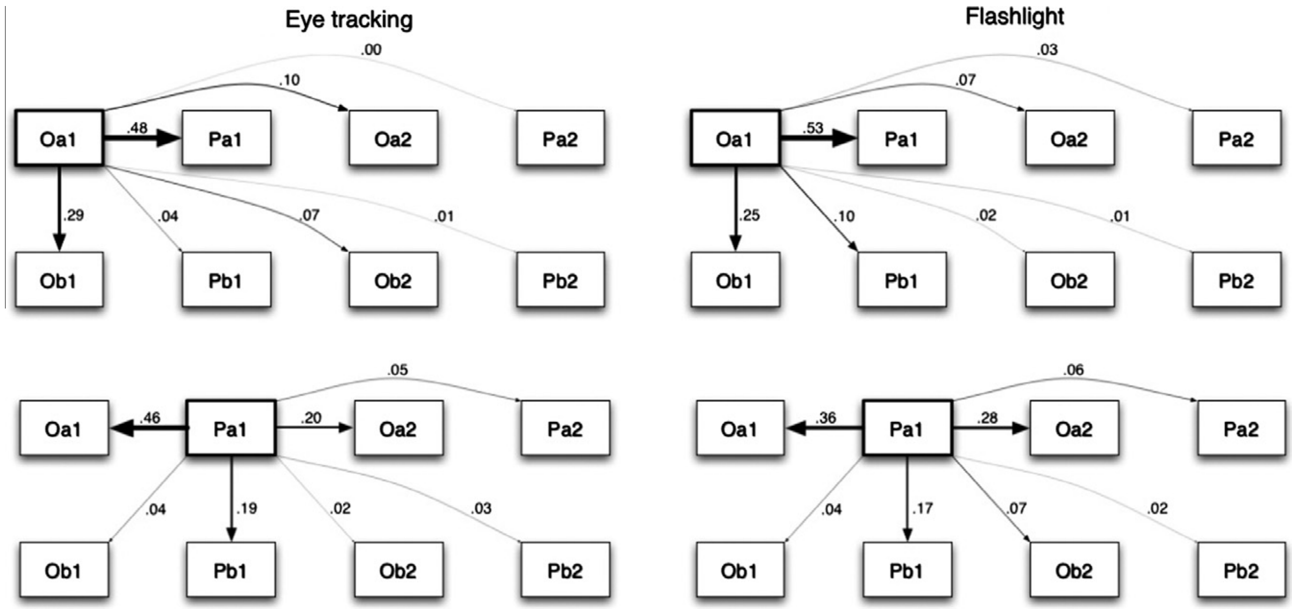

Fig. 6. Flashlight (right column) and eye-tracking (left column) transition figures for AOIs 1 (top row) and 2 (bottom row) in the choice task.

from a very easy way to allocate attention on a stimulus and also has the complete visual field of information always available. Flashlight reduces especially peripheral vision through the way the focus area presents information. This makes it harder to do large jumps from one side of a stimulus to the other. The coordination of eye + mouse movements will also add time in the information acquisition process for Flashlight.

Relating to the data we collected, we attribute the differences between the two methods to the following reasons: The outcome data might differ because of the different samples used in the experiment (see above). While the laboratory (eye-tracking) sample consisted of students from the local university, the online (Flashlight) sample was drawn from a large panel built from a broad set of participants. For example, the online sample was 8.5 years older on average and had a more evenly balanced gender distribution. The laboratory study clearly suffers from the fact that in a psychology department it is very hard to get an equal number of male and female participants due to the dominance of women studying psychology. There are three things to note at this point: (a) we do not make specific predictions for gender differences in our approach and also not expect a strong difference, however we wanted to point out the fact that our sample was unbalanced regarding gender, (b) in an online study one should expect a more balanced distribution and given the size of the panel easy ways to adjust to shifts into one direction, (c) individual differences within the samples might account for a certain part of the results. Regarding the different tasks types these differences may account for the slightly better adding results in the lab (with the student sample) and the reversed pattern for the reading task (better performance on the general knowledge questions for the non-students, however these differences were not significant). Additionally the general knowledge questions linked to the reading task might have been easier to answer for the non-students. Our intention in building the reading task was to generate a hard task in terms of process measures we reached this aim (see below) when inspecting the process data. In the gambles no such differences were found with both methods (and samples) following the predicted choice pattern.

Inspecting the data on the process level for task time and number of fixations several noteworthy patterns emerge. Clearly the number of fixations is larger in the eye-tracking recordings than in the Flashlight recordings regardless of the task at hand. This can be explained through the basic difference between acquiring information using ones eyes (a highly flexible, fast and well practiced) versus using the mouse (considerably slower and over time more strenuous). The mean fixation number also increased across the different tasks as predicted: the lowest number of fixations for the add task, more fixations in the choice task and the large amount of fixations in the read task. The results are not that clear for task time where longer completion times in the add and read task are in line with our above argument but a reverse (though not significant) pattern was found in the choice task with faster completion in FL. Additionally the predicted pattern of completion times (add < choice < read) was not found because the choice task resulted in the fastest responses. Different reasons can cause such a result: (a) Participants' motivation could have been higher in the add task due to the easy way to verify the correctness of their responses. In the choice task this verification is more complex and harder to see for the participant, hence lower motivation could cause shorter completion time. (b) There is a growing literature in decision making that emphasizes the usage of automatic processes in gambles (e.g., Glöckner & Betsch, 2008b). Accounting for such processes would change our predicted order in the sense that adding necessarily cannot rely on such processes and should result in longer task completion than a weighting process of an outcome probability pair which could be done automatically.

As a final step we looked into the patterns of acquisitions for the different tasks, a measure that gives us better understanding of how participants acquired information. The analysis of fixations per AOI in the choice task demonstrates that the interest to certain AOIs is similar regardless of the method used to record them (FL or ET). An even stronger argument can be generated when looking at the transition matrices for the choice task. Virtually the same overall transition probabilities and for all but one transition the same adjacent transition pairs emerge for FL and ET, despite the different samples and research locations.

5.3. Interaction of different dependent variables

The above described results offer the possibility to discuss an interesting situation: in the choice task the completion time data tell a different story than the fixation data (shorter completion time in FL but more fixations in ET for the same task). The solution to this somewhat contradictory result can be found in including the transition data (and to some extent the response data) to the analysis. As already mentioned the transition data show a highly similar picture for both methods. Therefore it can be concluded that our predicted process for the choices task of integrating the information (reading the numbers for a task first, multiplying the outcomes with the respective probabilities, adding the result up for each alternative and comparing the two alternatives) basically holds in terms of the acquisition but not in terms of the completion time.

5.4. Cross method comparison

In addition to what we discussed above regarding method comparisons with different samples another approach is to investigate the performance of data collected with one method to similar other methods. We will briefly compare our results with two papers investigating eye-tracking and Mouselab (Lohse & Johnson, 1996; Reisen, Hoffrage & Mast, 2008). Lohse and Johnson (1996) use different matrix sizes (2×2 and 7×7) with gambles (in a slightly different setup than in our study) as well as hypothetical choices between apartments. They found that it took participants longer to finish a task in Mouselab than in the eye-tracker while there were fewer fixations in Mouselab. Inspecting Fig. 3 (right half) we see the same patterns for Flashlight, e.g., in the add task where we used 5 AOIs compared to a 2×2 matrix in the Lohse et al. study. Even closer do the results of Reisen et al. (2008) resemble our own results. In their consumer choice task the authors found more fixations using eye-tracking (41.83) than Mouselab (22.35) a pattern also found in our study (see Fig. 3, right half).

Yet another way to compare and evaluate Flashlight is using the same set of stimuli between different methods. We decided to us a stimulus set with a long tradition in decision making research going back to Kahneman and Tversky (1979) in our choice task. This stimulus set was also more recently used in Johnson et al., 2008 in a process tracing study. While we do not want to go into the details of the discussion in this paper we want to point out that the process data collected by these authors with MouselabWeb (a within gamble search pattern, which hints at integration processes of the different attributes of a gamble) nicely fit to the results found in our study using both FL and ET. It can be concluded that with three different methods using three different samples a very similar picture in terms of transition matrices emerges. The participants in the different studies acquired information in a similar way providing further cross method evidence of the validity of our method.

5.5. Other applications

We studied tasks that allowed us to differentiate between easy and hard tasks both in reference to information acquisition as well as cognitive demand. Many other areas of application can be imagined for Flashlight. In a usability study we demonstrate ourselves that Flashlight offers easy ways to study comparisons between websites or computer programs. For questions from the consumer area one could imagine to study, e.g., food labels or other product relevant information. This could be done in different countries in an easy, convenient and cheap manner. For training in process tracing Flashlight could be used to build simple process tracing tasks quickly and analyze the results based on scripts (or additions to those) quickly.

The goal of Flashlight is not to offer a replacement for eye-tracking technology. We identified a niche in the currently available systems that we want to fill by providing an inexpensive, easy to administer, yet high-resolution measure of information acquisition processes in different task scenarios. Flashlight is opportune for researchers interested in studying the order and duration of information acquisition, but without the funding or apparatus to purchase eye-tracking equipment.

Acknowledgements

We want to thank the Faculty of Psychology (University of Bergen) Grant # 1751/1352 for providing funding for this research to the first author (Schulte-Mecklenbeck). The second author's (Murphy) contributions to this material are based upon work supported by the National Science Foundation under Grant No. (SES-0637151). This paper was presented as part of the special Society for Judgment and Decision Making symposium: 'Computer techniques in decision research' and at the Society for Computers in Psychology meeting in Chicago, November 2008. We want to thank Annika Boldt and Benjamin E. Hilbig for beta testing and comments, Simon Knaus for R support and three anonymous reviewers for comments that helped to improve the draft substantially.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chb.2011.03.004.

References

Abelson, R. P., & Levi, A. (1985). Decision making and decision theory. In G. Lindsay, & E. Aronson, (Eds.), *The handbook of social psychology* (3rd ed., pp. 231–309).

American Psychological Association (2009). Online psychology lab. <http://opl.apa.org/> Retrieved 19.06.09.

Babcock, J., Li, D., Parkhurst, D., & Winfield, D. (2010). Open eyes. <http://thirtysixthspan.com/openEyes/> Retrieved 08.02.10.

Bartlett, M. S. (1951). The frequency goodness of fit test for probability chains. *Mathematical Proceedings of the Cambridge Philosophical Society, 47*, 86–95.

Boldt, A., Schulte-Mecklenbeck, M., & Murphy, R. O. (2009). Flashlight: Online Erfassung von Blickbewegungen zur Evaluation von Webseiten [Poster]. Mensch und Computer. Berlin, Germany.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113*(2), 409–432.

Buscher, G., Dumais, S. T., & Cutrell, E. (2010). The good, the bad, and the random: An eye-tracking study of ad quality in web search. In *SIGIR '10: Proceeding of the 33rd international ACM SIGIR conference on research and development in information retrieval* (pp. 42–49).

Chen, M.-C., Anderson, J. R., & Sohn, M.-H. (2001). What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. In *Human factors in computing systems: Extended abstracts of CHI '01* (pp. 281–282). Seattle, WA: Association for Computing and Machinery.

Cutrell, E., & Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in Web Search. In *Proceedings of CHI'07, human factors in computing systems* (pp. 407–416). San Jose, April 2007.

Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455–470.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. London: Springer.

Edmonds, A., White, R. W., Morris, D., & Drucker, S. M. (2007). Instrumenting the dynamic web. *Journal of Web Engineering, 6*(3), 244–260.

Eger, N., Ball, L. J., Stevens, R., & Dodd, J. (2007). Cueing retrospective verbal reports in usability testing through eye-movement replay. In L. J. Ball, M. A. Sasse, C. Sas, T. C. Ormerod, A. Dix, & P. Bagnall, et al. (Eds.), *People and computers XXI – HCI... but not as we know it: Proceedings of HCI 2007*. Swindon: The British Computer Society.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241.

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137*, 673–690.

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology, 46*, 179–185.

Glöckner, A., & Betsch, T. (2008a). Do people make decisions under risk based on ignorance? An empirical test of the priority heuristic against cumulative prospect theory. *Organizational Behavior and Human Decision Processes, 107*, 75–95.

Glöckner, A., & Betsch, T. (2008b). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*, 1055–1075.

Gompel, R., Fischer, M., Murray, W., & Hill, R. (2007). *Eye-movements: A window on mind and brain*. Oxford: Elsevier.

Huber, O., Wider, R., & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica, 95*, 15–29.

Javal, E. (1878). Essai sur la physiologie de la lecture. *Annales d'Oculustique, 79*, 97–117.

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. (2008). Process models deserve process data. *Psychological Review, 115*, 263–272.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Krantz, J. H. (2008). Psychological research on the net. <http://psych.hanover.edu/research/exponnet.html> Retrieved 19.06.09.

Kühberger, A., Schulte-Mecklenbeck, M., & Ranyard, R. (2011). Windows for understanding the mind. In M. Schulte-Mecklenbeck, A. Kühberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research* (pp. 1–17). New York: Taylor & Francis.

Lohse, G., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes, 68*, 28–43.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Decision Processes, 16*, 366–387.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Reeder, R. W., Pirolli, P., Card, S. K. (2001). WebEyeMapper and WebLogger: Tools for analyzing eye tracking data collected in Web-use studies. In *Extended abstracts of the ACM SIGCHI conference on human factors in computing systems* (pp. 19–20).

Reips, U. (2008). Web experiment list. <http://genpsylab-wexlist.unizh.ch/> Retrieved 19.06.09.

Reips, U., & Lengler, R. (2005). The web experiment list: A web service for the recruitment of participants and archiving of Internet-based experiments. *Behavior Research Methods, 37*, 287–292.

Reisen, N., Hoffrage, U., & Mast, F. W. (2008). Identifying decision strategies in a consumer choice situation. *Judgment and Decision Making, 3*(8), 641–658.

Russo, J. E. (1978). Eye fixations can save the world: A critical evaluation and comparison with other information processing methodologies. In H. K. Hunt (Ed.). *Advances in consumer research* (Vol. 5, pp. 561–570). Ann Arbor, Michigan: Association for Consumer Research.

Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (Eds.). (2011). *A handbook of process tracing methods for decision research*. New York: Taylor & Francis.

Schulte-Mecklenbeck, M., & Neun, M. (2005). WebDiP – A tool for information search experiments on the World-Wide-Web. *Behavior Research Methods, 37*(2), 293–300.

Song, J. H., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision, 6*, 982–995.

Song, J. H., & Nakayama, K. (2008). Target selection in visual search as revealed by movement trajectories. *Vision Research, 48*, 853–861.

Spivey, M., & Dale, R. (2006). Continuous temporal dynamics in cognition. *Current Directions in Psychological Science, 15*, 207–211.

Spivey, M., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*(29), 10393–10398.

Tauscher, L. M., & Greenberg, S. (1997). How people revisit web pages: Empirical findings and implications for the design of history mechanisms. *International Journal of Human–Computer Studies, 47*, 94–137.

Todd, P., & Benbasat, I. (1987). Process tracing methods in decision support systems research: Exploring the black box. *MIS Quarterly, 11*(4), 493–512.

Willemsen, M., & Johnson, E. (2011). Visiting the decision factory: Observing cognition with Mouselabweb and other information acquisition methods. In M. Schulte-Mecklenbeck, A. Kühberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research* (pp. 21–42). New York: Taylor & Francis.