

# Philosophical Accounts of Causal Explanation and the Scientific Practice of Psychophysics

Tim C. Kietzmann  
NeuroBioPsychology Group  
Institute of Cognitive Science  
University of Osnabrück  
tkietzma@uni-osnabrueck.de

## Abstract

Philosophical accounts of causality and causal explanation can provide important guidelines for the experimental sciences and valid experimental setups. In addition to the obvious requirement of logic validity, however, the approaches must account for the generally accepted experimental practice to be truly useful. To investigate this important interconnection, the current paper evaluates different philosophical accounts of causation and causal explanation in the light of typical psychophysical experiments. In particular, eye-tracking setups will be used to evaluate Granger Causality, Probabilistic Accounts and Woodward's manipulationist approach. Upon coarse reading, the manipulationist perspective seems most suitable for a practical application, but there are manifold problems hidden in the details of the definitions. However, with some adjustments via standard tools of experimental design, these problems can be overcome and leave Woodward's account as the method of choice.

## 1 Introduction

In recent years, ever-advancing technological possibilities have allowed for more and more complex experimental setups. With this increase in complexity, the establishment of clear definitions and guidelines for what causal inferences can safely be drawn from these experiments is of crucial importance. One of the aims of the philosophical treatment of causality and causal explanation is exactly this: to provide clear-cut boundaries for legitimate causal inferences. However, in order to be helpful for experimental scientists, purely philosophical accounts of causal explanations need to be verified by the respective field of research and should not be naive to well-established and widely-acknowledged experimental practice. Apart from the obvious need for logical validity of these accounts, this verifiability is an important additional criterion of evaluation. This paper will take a first step into this direction by evaluating different approaches towards causal explanations with regard to the research field of psychophysics and overt visual attention (eye-tracking).

Having recently moved into the center of interest in vision research, the investigation of overt visual attention is particularly well suited for this endeavor because of its complexity and the many hidden effects that need to be accounted for. Shifts of overt visual attention are typically measured by eye-tracking, a technique that allows researchers to record eye-movements and secondary parameters, such as pupil dilation, blinking frequency, and saccade frequencies, with high temporal and spatial resolution. This way, eye-tracking enables us to know which information enters the visual system and thereby allows for a more detailed answer to the question of why it is selected in very natural and unconstrained settings. By applying cautiously designed experimental paradigms, researchers aim to differentiate between different theories not only of visual attention, but also of more general cognitive phenomena.

Before an evaluation of philosophical accounts of causal explanation can be performed, some basic descriptions of the utilized eye-tracking methodology need to be provided. Two typical analyses of eye-tracking data include Fixation Density Maps (FDMs) and measures of Feature-Fixation Correlations. The general idea of the first type of analysis is that it delivers a description of the average viewing behavior of the subjects. Normalized to a sum of one, the straightforward interpretation of these maps is that they describe the probability of fixations in different regions of the stimulus. Calculating these probability distributions for different experimental conditions then allows for objective measures of similarity, i.e. whether different conditions lead to significant differences in viewing behavior. Besides this spatial

analysis of fixation data, the analysis of fixated image features, such as color, luminance, or contrast information, can provide important additional information. It provides a description of what kind of image-properties were fixated, and hence which properties could have had a causal impact on the selection process. Here, a widely used technique involves the area under curve (AUC) of receiver operator characteristics. Put simply, it describes the relative importance of visual features at fixated positions by providing a measure of how well actual fixations, recorded from the subject in a selected condition, can be differentiated from randomly selected control fixations. The AUC of a feature is close to one if a feature is highly predictive of whether certain areas were fixated or not. AUC measures are therefore directly related to stimulus- or bottom-up-dependent influences on visual attention.

As mentioned before, the philosophical treatment of causality and causal explanations is extremely valuable with regard to the analysis and design of experimental paradigms, but it also needs to be in line with scientific practice. Since the two described methods, FDMs and AUC, are standardized and widely accepted measures, they provide a good starting point for the evaluation of the different philosophical accounts.

## 2 Accounts of Causality in the Light of Psychophysics

This section will focus on Granger causality, probabilistic accounts of causality and the manipulationist account of causal explanation. Although many more philosophical accounts of causality and causal explanation have been discussed in the classic and more recent literature, these three are of special interest because of their currently widespread application in the field of neuroscience or, as in the case of the manipulationist approach, because of their explicit focus on the direct applicability to experimental sciences. The discussion of each of the approaches will be started on more general grounds, before focusing on their interplay with eye-tracking setups.

### 2.1 Granger Causality

An overarching goal of many neuroscientific studies is to establish causal claims for different patterns of brain activity. Because in fMRI and PET experiments it is generally not possible to directly manipulate neuronal activity, findings are mostly bound to infer correlations instead of causation of which proof is truly sought. As a possible solution, many scientists apply the principle of Granger causality [1] to infer Granger causes from experimental data. The basic idea is that if the predictive power on measurement  $A_{t+1}$  is increased by adding prior measurements from a different region  $B_t$  to the priors of  $A$  ( $A_t$ ), then region  $B$  Granger causes activity in region  $A$ . Despite being straightforward and easily interpretable, it is not without problems. The first problem lies in the general validity of the inferred causal claims. In cases in which activity in  $A$  and  $B$  are both caused by a third process  $C$  with a certain additional time-lag  $\Delta t$  on  $B$  but not on  $A$ , Granger causality can be inferred although there is in fact no causal connection between  $A$  and  $B$ . This clearly poses a huge constraint onto potential interpretations of the data and possible causal inferences.

Apart from the above-mentioned methodological problems of Granger causality, further difficulties arise with regard to the general applicability to typical questions posed in eye-tracking studies. As mentioned before, the standard applications of Granger causality try to infer causal mechanisms between different brain regions, which typically cannot be subject to direct manipulations. In contrast to this, eye-tracking paradigms do not have the goal of inferring causality between different fixation sequences, which are the equivalent to activity in imaging techniques, but to reveal causal mechanisms between stimulus and task as independent variables and fixation patterns as dependent variables. Thus, a more suitable account of causality should not cover time-series, but a more direct matching between external causes and recorded effects.

### 2.2 Probabilistic Models

A more general treatment of causal explanations, which does not rely on the analysis of time-series, is the class of probabilistic accounts. Being directly related to Granger causality, the overall pattern of causal

reasoning is to infer that  $B$  causes  $A$  if  $B$  increases the probability of  $A$  (iff  $P(A) < P(A|B)$ ). More advanced accounts include the additional requirement that there exists no cause  $C$  such that  $P(A|C) = P(A|B, C)$  [2]. Although this extended version deals nicely with many of the problems of common causes, it does not solve the issue of causal preemption. As described by Woodward [3], if two gunmen shot at one person, both having a certain hitting probability ( $0 < P < 1$ ) and if, furthermore, only one of them hits the target and the person dies, then still both shots raised the overall probability of the person being killed and therefore both have to be interpreted as being the cause of the person's death. Since this is not the case, the causal inference is erroneous. Finally, probabilistic models typically do not exploit the temporal direction of the interaction between variables. This is clearly problematic for the resulting causal inferences.

Again, additional problems arise specifically in the light of typical eye-tracking analyses. The first problem is due to the interpretation of fixations as effect. Typical eye-tracking setups investigate whether patterns of fixations differ between conditions. How can this be modeled in a probabilistic framework? A possible solution would be to see the existence of a fixation as an effect for which the probability can be altered. Nevertheless, this admittedly naive approach is not applicable because fixations exist in all conditions. There are two alternatives to escape this: first, each pixel in the image could be interpreted as a potentially fixated region. An effect would then increase the probability of fixating this one pixel. Unfortunately, this leaves us with not one, but a large number of effects to be considered, although the experimental setup targets only one. The second option would be to regard patterns of fixations as effects. This way, the probability of a certain pattern would increase given the presence of a certain task. Although this approach seems reasonable, the problem lies in the definition of pattern-similarity. To determine significance, it has to be decided whether two patterns are equal and therefore count as the same effect. This clearly introduces unwanted subjective parameters into the analysis. In addition to the interpretive difficulties, the described solutions are counterintuitive with regard to standard experimental thinking. A more suitable approach would be to say that not the probability of the effect as such, but the probability distribution of the observed variable is altered in different conditions (as for instance expressed by the analysis of FDMs). According to this view, the investigated effect relates to changes in the underlying probability distribution of the effect-variable instead of the probability of the effect. Put differently, changes in fixation patterns are viewed as being the result of an alteration of the underlying probability distribution, which makes fixations at certain locations more probable than others. This approach overcomes the difficulties of the previous accounts and is close to the scientific practice. However, by relating effect probabilities to probabilities of changes in the underlying random variables, the probabilistic account is being redefined via manipulationist terms.

### 2.3 Manipulationist Accounts

There are numerous reasons why a manipulationist account of causal explanations is appealing with regard to experimental practice. First and foremost, changing experimental conditions by intervening on the independent variables and relating a notion of causality on changes in the dependent variable is close to the common principles of experimental design and scientific thinking. Moreover, causal claims established this way do not rely on the doubtful definition and presence of a natural law, as would for instance be required by Deductive-Nomological Model [4]. Another potential benefit is that this approach allows viewing causal reasoning and experimental findings as referring to "what-if-things-had-been-different" questions. If a wide variety of these questions can be answered, i.e. if many systematic changes can be explained by a model, then the overall behavior of and causality in the system can be said to be very well understood. A very prominent manipulationist account of causal explanation was put forward by Woodward [3]. In his book, an important claim is that the approach is particularly close to scientific procedures and conventions. To assess the validity of this claim, we will first take a rather general view on typical experimental paradigms in eye-tracking and psychophysics. Following this, a more detailed analysis of the underlying definitions will be provided.

The most fundamental notions of Woodward's approach are direct and contributing causes<sup>1</sup>.

(DC) A necessary and sufficient condition for  $X$  to be a direct cause of  $Y$  with respect to some variable set  $V$  is that there be a possible intervention on  $X$  that will change  $Y$  (or the

---

<sup>1</sup> Definitions taken from [3]

probability distribution of  $Y$ ) when all other variables in  $V$  besides  $X$  and  $Y$  are held fixed at some value by intervention.

(NC\*) If  $X$  is a contributing type-level cause of  $Y$  with respect to the variable set  $V$ , then there is a directed path from  $X$  to  $Y$  such that each link in this path is a direct causal relationship; [...] if  $X$  causes  $Y$ , then  $X$  must either be a direct cause of  $Y$  or there must be a causal chain, each link of which involves a relationship of direct causation, extending from  $X$  to  $Y$ .  
(p. 55ff)

That is, a direct cause can be established if a change in  $Y$  can be brought about via a manipulation of the state of  $X$  while all other variables are held fixed. A contributing cause, by contrast, describes a sequence of direct causes. In both cases  $X$  can be said to cause  $Y$ . Both definitions rely on the notion of a "surgical" intervention, which is of central importance in Woodward's account. This is not only because it required for DC and NC\* but also because a clear understanding of a proper intervention is crucial for the success of an experimental design and to reveal potential sources of errors.

(IV)  
I<sub>1</sub>.  $I$  causes  $X$ .  
I<sub>2</sub>.  $I$  acts as switch for all the other variables that cause  $X$ . ...  
I<sub>3</sub>. Any directed path from  $I$  to  $Y$  goes through  $X$  ...  
I<sub>4</sub>.  $I$  is (statistically) independent of any variable  $Z$  that causes  $Y$  and that is on a directed path that does not go through  $X$ .

(IN)  $I$ 's assuming some value  $I = z_i$  is an intervention on  $X$  with respect to  $Y$  if and only if  $I$  is an intervention variable for  $X$  with respect to  $Y$  and  $I = z_i$  is an actual cause of the value taken by  $X$ .  
(p. 99)

In essence, an intervention ( $I$ ) and therefore the drawn conclusions are valid if  $I$  is the only causal factor that determines the state of the independent variable ( $X$ ) and there is no causal link between  $I$  and the dependent variable ( $Y$ ) that does not pass through  $X$ . The latter is required to make sure that the causal inference about  $X$  and  $Y$  is truly based on the effects of  $X$ .

A main problem of the aforementioned approaches is the inability to account for the analysis of fixation data (FDMs). As can be seen from the definition of direct and contributing cause, Woodward defines changes of dependent variables as changes in their underlying probability distribution. This directly matches the analysis of FDMs, since different FDMs can be interpreted as the result of changes in the underlying probability distribution of the effect variable. Furthermore, Woodward's definition of an intervention matches the assignment of subjects to different conditions in the experimental setup. Finally, Woodward's approach accounts for different levels of explanatory depth. Invariant generalizations, which can answer a greater range of what-if-things-had-been-different questions, are expected to exhibit an increased explanatory power. This provides researchers with an elegant tool with which they can compare the generality of models and approaches. Transferred to the domain of visual attention, more general models should continue to hold under a range of different tasks and stimuli (interventions on the stimulus and top-down variables) and explain a large part of the variance in the data.

Summing up, Woodward's approach towards causal explanations seems to be very well equipped to cover the scientific practice in the research on visual attention. The strong emphasis on changes in dependent and independent variables, the definitions of direct and contributing causes and the notion of invariance have direct equivalents in typical eye-tracking paradigms. Finally, the provided definitions of proper interventions can be used as clear guidelines for the design of experimental paradigms and help clarifying the actual causal mechanisms in the experimental setup. Given the potential benefits of Woodward's approach, a more detailed look at his definitions is required to assess, whether it is indeed in agreement with the scientific practice. For this purpose, the experimental setup of a recent eye-tracking study by Betz et al. [5] will be used.

### 3 A closer look at Woodward's approach: An eye-tracking Case Study

A good part of research on overt visual attention tries to assess the exact mechanisms, which underlie the selection of fixation targets. A widely-examined theory suggests that the direction of gaze is determined by bottom-up, stimulus-dependent features in the visual world. This account is based on the hypothesis that every visual feature is represented in a feature-map. These maps are then combined into an overall representation of saliency - the saliency map [6], in which high values correspond to high 'interestingness' and to potential targets of eye-movements. At the same time, however, it is well established that different tasks can alter our viewing behavior in a systematic, top-down directed manner. If we agree that image features and higher level processes can both influence fixation behavior, the resulting question is whether and in which way the two interact. The weak top-down hypothesis suggests that the viewing behavior in different tasks can be explained by an altered weighting of the importance of different features. For instance, if some tasks put more emphasis on the color-channels, the resulting saliency map would change accordingly, emphasizing colored regions. Contrary to this, the strong top-down hypothesis suggests that fixation behavior can be changed independently from the features of the bottom-up system. That is, certain regions of the stimulus are emphasized independently of the occurring features (a possible realization of this account would be a spatial bias). To shed light onto this issue, an eye-tracking experiment was designed to test these two opposing hypotheses [5]. The experimental idea was to introduce different top-down processes by defining different tasks, which the subjects have to perform while viewing the stimuli. To check for an overall change in viewing behavior, FDMs are calculated, as described before, and then checked for statistically significant differences. To study task-related changes in the bottom-up saliency, stimulus features at fixation-points are evaluated based on AUC values. Again, AUC provides a measure of how predictive a certain feature is for the distinction between actual- and control-fixations. The study had two potential outcomes. If different viewing behavior was found in different tasks, accompanied by significant changes in the AUCs, this would speak in favor of weak top-down. In case of strong top-down, the task-dependent processing is expected to be independent of the bottom-up system and changes in the fixation behavior should therefore not be correlated with changes in the bottom-up system, as again revealed by the AUC values.

Figure 1 shows the graph of the experimental setup. The top-down variable is intervened upon with different tasks, and the effect variable is the recorded fixation behavior. In the case of strong top-down, the different tasks should act as direct cause in fixation selection. In the case of weak top-down, they need to be interpreted as contributing causes, because they alter the viewing behavior via changes in the bottom-up feature weighting. Despite its simplicity and validity, this paradigm is not without problems when being interpreted in the light of Woodward's definitions.

#### 3.1 Problem 1: The problem of background effects

According to Woodward's definition, a proper causal claim can only be established when all background variables that are not in the causal chain, but which could affect the effect variable, are held fixed. With regard to the currently described setup, this means that all other variables, which could have an effect on the fixation behavior need to be fixed. This includes variables like 'gender', 'mood', or 'age', which are all candidates to influence viewing behavior. A straightforward solution would be the use of a within-subject design. If the testing of different conditions is kept within one well-controlled subject, the external effects can be minimized. Unfortunately, this solution is not applicable in the current case, because it would add the requirement to repeatedly use the same stimuli in the different conditions to control for bottom-up stimulus effects. This introduces uncontrollable memory effects (see below).

#### 3.2 Problem 2: Controlling the independent variable

The second problem is directly related to Woodward's definition of a proper intervention. A correct intervention has to be surgical, i.e. it is required to 'switch off' all other variables with a causal impact on  $X$ . When being considered in the current experimental setup, it is thus required that the intervention acts as a switch for all other causes of the top-down process apart from the task-variable. Unfortunately, potential changes in the top-down variable can be induced by prior knowledge, task-sequence effects, stimulus-

sequence effects, and memory effects (to name just a few). If the prior knowledge in one subject differed from the knowledge of another, and this is not the actual intervention, then an effect found later could simply be due to this difference. Similarly, if the order of stimuli and tasks is equal for every subject, it is again not possible to attribute changes in the dependent variable to the manipulation on the independent variable because the actual cause might have been the fixed sequence. Clearly, it is not possible to bypass these issues in an individual experimental run. The problem of memory-effects becomes especially tricky together with Problem 1. As a solution to the problem of background effects on the dependent variable, a within-subject design was proposed. However, this approach requires the experimental setup to show the same images repeatedly in different conditions. This results in uncontrollable memory-effects, which can now influence the independent variable (top-down). Thus, despite solving Problem 1, this approach makes it impossible to meet the criteria for a 'surgical' intervention.

### 3.3 Problem 3: Variables that cannot be fixed

Problem 1 and 2 are related to background variables and the potential problems of 'surgical' interventions. That is, the problems affect both the 'effect' and the 'cause' variable. The main idea of the experiment was to establish whether top-down differences act as direct or indirect cause. For this, a correct 'manipulationist' experiment would require fixing the intermediate variable, the bottom-up features at fixation points, and then to investigate whether there are still differences in viewing behavior in different conditions. If such differences can be found, they can be assumed to be the direct effect of the different tasks and not due to changes in the bottom-up features (as predicted by the strong top-down hypothesis). Unfortunately, fixing the bottom-up feature correlations would require fixing the subject's eye-movements, which clearly runs counter to the overall idea of an unrestrained eye-tracking setup. Thus it seems, in the light of Woodward's definitions, not only that the requirements for effect- and cause- variables clash with the experimental setup, but also that the overall design of the experiment is flawed because it cannot establish any evidence for a direct versus an indirect causal effect.

### 3.4 Solutions

Although the described experiment is in line with the scientific practice in the field of psychophysics and eye-tracking, and although its causal claims are accepted as being valid, it conflicts with Woodward's very basic definitions. The question is therefore which experimental procedures can be used to bring the two together.

Problem 1 could be solved by using a within-subject design. This, however, leads to the problem of memory effects (Problem 2). A solution to both problems is given by across-subject designs. In this case, every stimulus is shown only once per subject and condition and the later analyses are calculated across subjects. Although the stronger claim of an individual change is not supported, finding an effect across subjects shows that, in the mean, the effect is present (as in this case different viewing behavior). This claim is widely regarded as equally valid. In addition to canceling out background effects on the effect variable and preventing memory effects, this approach also excludes differences in prior knowledge that could influence the top-down variable.

Problem 2 stated furthermore, that sequences of stimuli and tasks could potentially influence the top-down variable in an uncontrolled manner. This problem can be overcome by randomizing the task and stimulus sequence across subjects, as well as the mapping of stimuli to conditions. The general reasoning of randomized across-subject designs is that all effects, which cannot be directly controlled are kept stable across conditions because the effects of these variables average out. In short, this approach renders the different experimental groups equivalent such that the measured changes on the effect variable reflect only the experimental manipulation.

Problem 3 dealt with the overall setup of the experiment. A clean manipulationist approach would be to fix the effects of different feature-weighting and then to check for different viewing behavior in different tasks. Because this is simply not possible, the question is how it can be established that a variable is nevertheless stable across conditions. One possibility would be a simple observation of the variable. If an observation of the complete state of the variable is possible, and it remains fixed without external control,

then there seems to be no argument against the validity of this procedure. In theory, an observation of a (stable) variable should thus suffice to fulfill the requirement of being fixed independently of the means by which this is accomplished. Unfortunately, the requirement of full observability cannot be met in the current case because this would imply that all possible bottom-up feature-types need to be observed. This is impossible, because there exists no exhaustive list of feature-representations in the lower levels of the human visual system. Furthermore, continuous variables could change in ways that are not covered by the accuracy of the recording device. These changes would also stay unnoticed. Statistical significance is also no solution, because finding no significant differences in the different conditions does not imply that there is no effect. As a solution to this problem, the authors of the case study discussed above applied computational modeling to show that the differences found in bottom-up feature weightings were not sufficient to explain the observed differences in viewing behavior. Therefore, most of the causal impact can be attributed to the direct causal route from the top-down task to viewing behavior.

The final causal graph is shown in Figure 1. By across-subject experimental design, effects of background and hidden variables are held fixed, and memory effects are prevented. When combined with randomization, task- and stimulus-sequence effects can be prevented. Variables, which cannot be fixed, can be assumed to be stable if the complete state of the variable can continuously be observed or if computational modeling can be used to assess the potential impact of the variable in question.

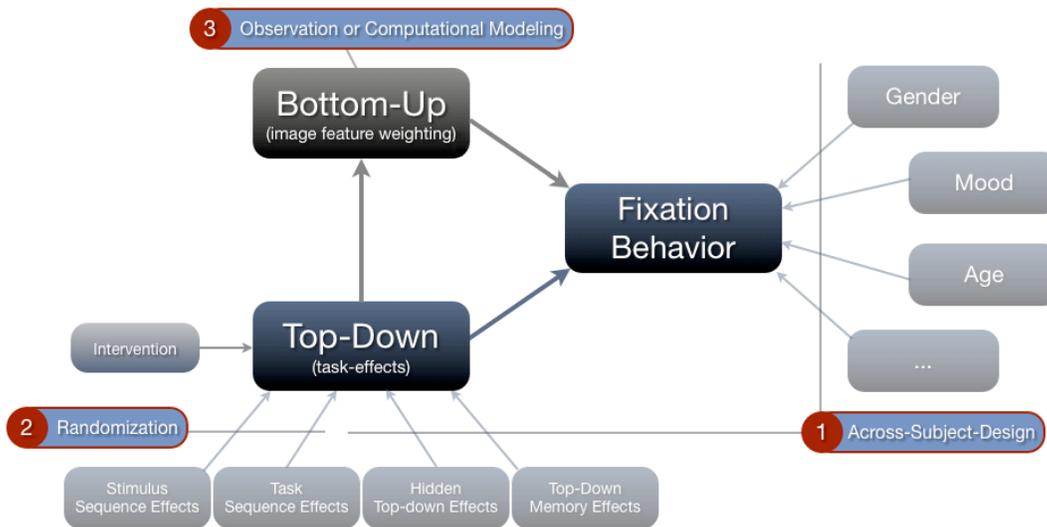


Figure 1: The experimental setup with manipulated, observed, and monitored/modeled variables. The independent variable, the task, was intervened upon to check for the influence of different top-down processes on fixation behavior. By applying an across-subject design, unwanted background effects on the cause- and effect-variables cancel out. Sequence effects can be prevented by randomization. Because the bottom-up features are dependent on the final fixations of the subjects, they cannot directly be controlled (and therefore not be fixed). In order to check whether they were stable across different tasks, which would be in line with the strong but not the weak top-down hypothesis, they were used for parameter estimation in a computational modeling approach.

## 4 Discussion

In this paper, it was argued that philosophical accounts of causation and causal explanation can provide highly beneficial guidelines for the experimental sciences. However, this relation is not a one-way route, because the philosophical ideas have to agree with well-established scientific practices and reasoning. Using the case of psychophysics and eye-tracking paradigms, it was shown that Granger-causality and probabilistic approaches do not only exhibit problems regarding the general establishment of causal

claims, but also because they are not in line with the type of reasoning and data usually present in eye-tracking studies. Following this, Woodward's manipulationist approach was discussed. Despite seeming like a well-suited candidate on a rather general treatment, a strict reading of Woodward's definitions of direct cause and proper interventions shows that they are in clear conflict with standard experimental conventions. This was exemplified by the analysis of a case study.

Fortunately, these problems can be overcome by very basic control mechanisms of statistics and experimental design: across-subject design, randomization and computational modeling/observations. Without actually fixing the potentially confounding variables, as explicitly required by Woodward, it can still safely be assumed that their effects cancel out and that the established causal claims remain valid. Thus, with the currently described modifications, Woodward's account is in line with the scientific practice in psychophysics and can provide a very powerful tool for the experimental design and the analyses of causal inferences not only in eye-tracking setups.

## 5 Acknowledgements

The author wants to thank Achim Stephan and Vera Hoffmann-Kolss for their helpful comments on an earlier version of this manuscript.

## References

- [1] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Essays in Econometrics: Collected Papers of Clive WJ Granger*, 2001.
- [2] W. Salmon, *Scientific explanation and the causal structure of the world*. Princeton University Press Princeton, NJ, 1984.
- [3] J. Woodward, *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, USA, 2005.
- [4] C. Hempel and P. Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Technology: The Technological Condition-An Anthology*, vol. 135, p. 75, 2003.
- [5] T. Betz, T. Kietzmann, N. Wilming, and P. König, "Investigating task-dependent top-down effects on overt visual attention," *Journal of Vision*, 2010.
- [6] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.