DETECTING MINIMAL DIFFERENCES IN THE OBSERVATION OF EVIDENCE IN INQUIRY LEARNING

Carolin Richtering
University of Twente

Abstract

Evidence evaluation is one key skill in inquiry learning. Chinn and Malhotra (2002) found however, that evidence evaluation can be severely impeded because learners have problems at making good observations. This is especially the case when the evidence is ambiguous. The present study investigates why the observation of evidence is impeded and thereby draws on existent knowledge and methods from psychophysiology research. An inquiry specific task is set up where minimal temporal differences can be manipulated. It is assessed whether a difference between the observation of minimal differences in the visual and auditory modality exists and whether cross-modal interaction effects and (the strength of) prior beliefs influence performance. The results show that typical psychophysical results could be replicated. Attending to auditory stimuli or both auditory and visual stimuli yielded threshold values of about 80 ms. Attending to visual stimuli only resulted in a slightly better performance, especially at small temporal differences. This was not expected from research and might be explained by the inquiry specific characteristics of the visual stimuli. No cross-modal interaction effects were found. Prior beliefs biased results in such a way that a correct predictions led to better observations whereas incorrect predictions had no effect. Being more certain intensified this effect. Suggestions for supporting learners in observing evidence are given.

Samenvatting

The evalueren van evidentie is een sleutelvaardigheid in onderzoekend leren. Chinn en Malhotra (2002) hebben echter ontdekt dat het evalueren van evidentie sterk belemmerd wordt omdat leerlingen problemen hebben goede observaties te maken. Dit is in het bijzonder het geval als de evidentie ambigu is. De voorliggende studie verkend waarom het observeren van evidente belemmerd wordt en maakt daarbij gebruik van bestaande kennis en methoden van psychofysiologisch onderzoek. Er werd een taak opgesteld die kenmerkende eigenschappen van onderzoekend leren bevat en waar heel kleine tijdelijke verschillen gemanipuleerd kunnen worden. Het werd onderzocht of tijdens de observatie van heel kleine verschillen tussen de visuele en auditieve modaliteit bestaan and of interactie-effecten tussen de modaliteiten en and (de sterkte) van voorafgaande overtuigingen de prestatie beïnvloeden. De uitkomsten laten blijken dat kenmerkende psychofysiologische resultaten gerepliceerd konden worden. Aandacht voor enkel auditieve stimuli of auditieve en visuele stimuli heeft tot drempelwaarden van 80 ms geleidt. Aandacht voor enkel visuele stimuli resulteerde in iets betere prestaties, in het bijzonder bij kleine tijdelijke verschillen. Dit werd opgrond van onderzoek niet verwacht en zou door het voor onderzoeken leren kenmerkende karakter van de visuele stimuli verklaard kunnen worden. Interactie-effecten tussen modaliteiten konden niet worden geobserveerd. Voorafgaande overtuigingen hebben de resultaten op een degelijke manier beïnvloed dat correcte voorspellingen tot betere observaties leidden, terwijl incorrecte voorspellingen geen effect hadden. Een grotere zekerheid over de voorspelling heeft dit effect versterkt. Aanbevelingen voor de ondersteuning van leerlingen bij het observeren van evidente worden gegeven.

The goal of science education in the 21st century is not to have the “knowledge of a lot of science, but rather the understanding of how science really works” (Durant, 1994, p.83). It is thus not important to recall a large body of scientific facts, but to understand how science works and how it is based on the analysis and interpretation of evidence. This tenet is in line with the pedagogical approach of inquiry learning. Inquiry learning promotes the understanding of science, because inquiry learning is learning science by doing science. The core activities that take place during inquiry learning can be explained by the dual space search theory from Klahr and Dunbar (1988). Klahr and Dunbar see the inquiry process as a search in two related problem spaces: the hypothesis space and the experiment space. The learner first performs a search in the hypothesis space and formulates (partial) hypotheses based on prior knowledge. The subsequent search in the experiment space is constrained by these hypotheses. The learner tries to find an experiment that can discriminate among rival hypotheses. As a final step, he evaluates the evidence.

The process of evidence evaluation, which was somewhat neglected in the initial discussion of the dual space theory, was elaborated in later work (Klahr, Fay & Dunbar, 1993). This additional focus originates from the work of Kuhn (1989) who states that the skills at differentiating and coordinating theory and evidence lies at the heart of scientific thinking. This is in line with Zimmerman’s (2007), who argues that evidence evaluation is one of the key skills for promoting conceptual change and scientific understanding (Zimmerman, 2007). Evaluating evidence in inquiry learning demands from the learner to collect data through observation, make interpretations and adapt their hypotheses accordingly (Chinn & Malhotra, 2002).

Especially novices, however, seem to have great difficulties at this stage of the inquiry process. Chinn and Malhotra (2002), for instance, found that not the interpretation of results or drawing of conclusions is causing serious troubles, but that evidence evaluation is most severely impeded during observation. In their study, Chinn and Malhotra let simultaneously drop two rocks with approximately the same size but different weights. The outcome of the rock-dropping experiment was counterintuitive for most children, because they expect that heavy rocks fall faster than lighter ones. Children in 4th grade do not have an (elaborate) idea about gravity yet. Chinn and Malhotra (2002) found that children who had incorrect prior beliefs, however, were not predestined to make observations that would fit their predictions. They conclude that evidence observation was impeded in their experiment because only 44% of incorrect predictions were followed by correct observations. Children who had “correct” prior beliefs, however, were more likely to make the right observation. This asymmetrical bias could be confirmed in Chinn and Malhotra's other experiments of the same research. In line with the findings of Chinn and Malhotra, Klahr and colleagues (1993) observed that bias in favor of a particular hypothesis may affect which aspects of evidence are attended to and encoded.

Chinn and Malhotra (2002), however, did not address the issue of whether very small visual and auditory differences can be detected in the ambiguous experiment. When two rocks are dropped simultaneously, they are difficult to observe. Chinn and Malhotra propose that that listening might be more effective than just watching. But scaffolding prompts that were thought to facilitate observation ('not only focus on the falling objects but also listen carefully'), did not improve the children's performance.

In order to better understand why inquiry learners often find it difficult to observe evidence, especially when the evidence is rather ambiguous, it is crucial to understand what happens during this process and where the pitfalls are. A substantial body of evidence focusing on this ability exists in the field of psychophysiology and will therefore be consulted. Psychophysical research investigates the relationship between physical and hence quantitatively measurable stimuli and the way human beings experience these stimuli.

Investigating evidence observation in the framework of psychophysiology

A very important skill for observing evidence is the ability to correctly judge minimal temporal differences. In psychophysical research, temporal order judgments (TOJ) tasks are used as a measure for the temporal resolution of the visual and auditory system. On a typical TOJ task, participants are presented with two consecutively appearing sensory stimuli, separated by a very brief inter-stimulus interval (ISI) and then asked to decide which of the stimuli they perceived first (Sternberg & Knoll, 1973).

A main assumption in psychophysical research is that there is no single point where a temporal order suddenly becomes detectable. Instead there are a range of intervals where participants can sometimes report the correct temporal order and sometimes not (Moore, 1982). The percentage of correctly observed TOJs can be related to the length of the ISI in a psychometric function. Psychometric functions assume a sigmoid form because very short inter-stimulus intervals will never be identified correctly whereas clearly perceptible ISIs always will. The temporal order threshold (TOT) marks the shortest time interval between two sensory events necessary for a person to be able to identify the correct temporal order.

Various researchers have tried to prove that temporal perception occurs quite similarly in all sensory modalities (Pöppel, 1997; Sternberg & Knoll, 1973; Szelag, Kanabus, Kolodziejczyk, Kowalska & Szuchnik, 2004). They postulate that there is some kind of central mechanism underlying this ability. Although this assumption has been substantiated by some compelling results, other studies report different temporal order thresholds for the auditory and visual modality. Vroomen, Keetels, de Gelder & Bertelson (2004), for instance, claim that sensory modalities differ in physical transmission and neural processing and conclude therefore that TOT values in the visual and auditory modality differ. Jaśkowski, Jaroszyk and Hojan-Jezierska (1990) found a relative latency of the visual modality compared to the auditory modality of about 30 ms. Aschersleben and Müsseler (1999) performed a comparable study and estimated this difference to be approximately 15 ms. Temporal perception is coded more accurately in the auditory modality, whereas spatial perception is coded more accurately in the visual modality (Binda, Morrone & Burr, 2010; Wada, Kitagawa & Noguchi, 2003).

When both visual and auditory stimuli are present, cross-modal interaction effects are likely to occur. Binda and colleagues (2010), for example, found that sound information was attributed more weight, when visual stimuli are blurred or presented during saccades (during this rapid ballistic eye movement brief visual stimuli get severely dislocated). Auditory stimuli then change the apparent timing of the visual stimuli. In a study from Wada and colleagues (2003), participants had to judge the change in the frequency of very brief visual and auditory stimuli. They found in visual as well as auditory TOJ tasks that the influence of the background modality that has to be ignored is stronger when the stimuli in question are ambiguous. These two examples of cross-modal interaction show, that in cases of ambiguity, one modality can influence the apparent timing of the other. Lower threshold values due to cross-modal interaction effects can be expected in such cases.

Different neuronal mechanisms are involved in temporal processing, depending on the characteristics of the stimulus (Ulbrich, Churan, Fink & Wittmann, 2009), such as for example the tone frequency and the complexity of the signal. For high tone frequencies of 8 kHz, relatively small threshold values of about 4.6 ms were found in a gap detection task is (Florentine, Buss & Geng, 1999). These tasks are similar to TOJ tasks, except that participants just have to distinguish whether there is an interval between stimuli and not what the TO of the stimuli is. With decreasing frequency, thresholds values increase. At 1 kHz for example, the threshold value reaches 12.7 ms and at 0.25 kHz ultimately 88.1 ms. The human information processing system is able to detect temporal order difference for non-complex stimuli (e.g. two tones of different frequencies) at thresholds as low as 10 ms (Fink, Churan &

Wittmann, 2005). For complex stimuli such as colors and stimuli with different orientation, mean temporal order thresholds up to 120 ms were found (Ulbrich et al., 2009).

Individual differences generally have a rather small influence on the results of temporal order judgments. The level of practice, for example, does not seem to matter. In several studies with inexperienced as well as experienced participants, no significant differences were detected between these two groups (see, for example, Formby, Gerber, Sherlock & Magder, 1998). Gender differences are also rarely found in TOJ tasks and if present, the occurrence is restricted to stimuli with a simple click sound or with different positions (Ulbrich et al., 2009). The age of the participants only seems to matter below or above a critical age. Szelag and colleagues (2004) did not find any significant difference between the performance of young (19 - 25 years) and old (65 - 67 years) adults. However, very high TOT-values were found for centenarians. Szymaszek and colleagues (2006), in contrast, did not find differences between older (60 - 69 years) and younger participants (20 - 28 years). They propose that age-related deterioration depends on the physical characteristics of presented stimuli and of slower information processing. TOJ tasks are not suited for children younger than 7, because they have problems with understanding and executing the task (Bergwanger, Wittmann, von Steinbüchel & von Suchodoletz, 2004). Young children also seem to have more problems in detecting and discriminating between sounds. Although their hearing has matured, some yet undefined central process is not (Werner, 1996). By the age of 11 or 12, however, temporal resolution reaches adult values (Irwin, Ball, Kay, Stillman & Rosser, 1985).

The role of prior knowledge and beliefs has emerged as an important influence in evidence observation and evaluation (Zimmerman, 2007). The asymmetrical bias found in Chinn and Malhotra's (2002) series of experiments is a good example for this. Zimmerman furthermore proposes that also the strength of prior beliefs might influence evidence evaluation. Comparable distorting effects also prevail at evidence observation in TOJ tasks. Fernbach, Linson-Gentry and Sloman (2007), for instance, found that beliefs about causation can influence the perception of temporal orders. Participants in their study had to figure out the temporal order in which three sliders move. A TOJ task with another, yet unrelated temporal order followed. The study revealed that participants with incorrect judgments showed a bias in favor of the temporal order learned before. However, when asked, participants were not aware of this. Fernbach and colleagues (2007) propose that people use prior beliefs when they are uncertain about what happened. Events that were not attended to, were filled-in with prior beliefs.

Present study

This present study will explore the detection of minimal differences in the observation of evidence in inquiry learning. It will investigate whether the perception of minimal differences in evidence observation differs between the visual and auditory modality and what the influence of cross-modal interaction and (the strength of) prior beliefs is.

Psychophysiological research gives a general idea of what to expect when examining the observation of evidence more thoroughly and why observation might be impeded. It can be expected that, when examining evidence observation, typical results of TOJ tasks can be replicated. Various studies have shown that auditory stimuli are usually superior in temporal processing and therefore lower threshold values can be anticipated in the auditory variant. Furthermore, cross-modal interaction effects are common in psychophysical tasks and a positive distortion of TOJ performance scores at the presence of both visual and auditory stimuli, compared to only one type of stimuli, is likely to occur. The study from Fernbach and colleagues (2007) has very compellingly shown how even unrelated events can influence TOJs. It is therefore expected that the prior beliefs of participants influence the subsequent TOJ. It is, however, not yet assessed whether and how the strength of those prior beliefs influences the subjective experience of minimal differences.

The experimental design will follow the psychophysical research paradigm, augmented with an inquiry specific task. The experimental design will allow the manipulation of minimal differences, makes it possible that the visual and the auditory modality can be assessed apart and that cross-modal interaction effects and (the strength) of bias can be sufficiently examined. The previous discussion of psychophysical research has shown that stimuli characteristics influence the performance on TOJ tasks. It is therefore chosen to use auditory stimuli that are comparable to sounds in inquiry learning tasks, in terms of tone frequency, complexity and other stimuli characteristics. This means for the visual stimuli that they will not suddenly appear, as it is common TOJ task, but will be visible before the TOJ event happens. This was, for example, also the case in the rock-dropping experiment. Observer characteristics do not have to be taken into account for the choice of participants.

Method

Participants

Twentey-eight students (5 males, 23 females) participated in the study for course credits. They ranged in age from 18 to 25 years, with a mean age of 20.8 years (*SD* = 1.8). Participants reported normal hearing and seeing or wore glasses or lenses.

Materials

The experiment was performed with a computer simulation of two cars driving down differently shaped streets and crashing against a wall at the end of these streets. The temporal order (TO) in which the cars crash and the interval between the two crashes, the inter-stimulus interval (ISI), could be manipulated. There were three different variants of the experiment, presentation of visual stimuli only, presentation of auditory stimuli only and presentation of both visual and auditory stimuli.

The presentation of the stimuli took place according to the method of constant stimuli. With this method, the step level in which the temporal differences of the two stimuli increase is set beforehand (Dai, 1995). ISI values were 20, 40, 60, 80, 100 and 120 ms. The two extreme ISIs, 20 and 120 ms, were presented 5 times and the remaining intervals 10 times within each variant. The advantage of this method is that every participant received the same TOJ-trials which helped to assess the influence of cross-modal interaction and prior beliefs. The participants are provided with a 2AFC (two answers forced choice) response format with an additional category for uncertainty (“the cars crashed simultaneously”). Adding a response option for uncertainty makes participants feel more at ease because they are not forced to choose (Kaernbach, 2001; Klein, 2001).

The visual stimuli consisted of a presentation of two different streets, with a yellow and a green car driving down these streets and crashing against a wall. The two streets were chosen from a pool of a 20 streets which in turn were made up of five differently shaped streets of approximately the same length (567.63, 569.77, 569.87, 570.10 and 574.88 pixels) in a normal, vertically mirrored, horizontal mirrored and vertical and horizontal mirrored version (see Figure 1). A card-sorting task prior to the conduction of the experiment revealed that participants could not simply estimate the length differences, but rather differed in their judgments. Both cars always drove with the same speed, which depended on the ISI that had to be achieved. The speed ranged between 2 and 13 pixel/second.

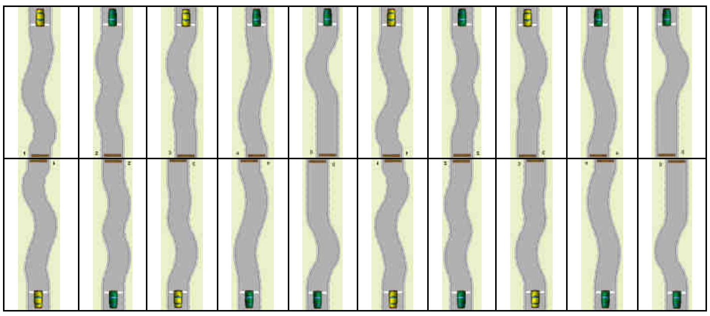

Figure 1. Pool of differently shaped streets with approximately the same length, which form the basis of the visual stimuli for the temporal order judgment (TOJ).

The auditory stimuli were presented via a circumaural headphone (Sennheiser PC 151). A pleasant volume was chosen by the participants before the experiment started and stayed the same throughout the session. The crashing sound of the cars was a sudden burst sound with a Gaussian sound envelope reaching 5455 Hz.

In the visual variant only the visual stimuli were presented; no crash sound was played. In the auditory variant, the streets were not visible but hidden in fog so that participants had to attend to the auditory stimuli only. In the visual & auditory variant, both visual and auditory stimuli were presented.

The experiment was performed in a normally lit room. Participants sat at 60 cm distance from a 15 inch screen with a resolution of 1024 x 768 pixels. The room was not soundproof but was situated in an office track and background noise therefore remained very low. Furthermore, noise was additionally attenuated by the headphones worn by the participants.

Prior beliefs were measured before each trial with a three-response format (right car will crash first, left car will crash first, both cars will crash simultaneously). The strength of prior beliefs was indicated on a 5-point Likert scale ranging from very uncertain to very certain. In the auditory only variant, the streets could be made visible for the prediction by pressing a button that made the fog disappear (the fog reappeared when the button was released). The temporal order judgment was assessed in the same three-response format as the prior beliefs. No feedback was given.

Design

The experiment had two within-subjects factors, variant (visual only, auditory only and visual & auditory) and inter-stimulus interval (20, 40, 60, 80, 100 and 120 ms). The chronological order of variant presentation was counterbalanced over participants. The temporal order in which the cars crashed, left first or right first, was randomized over the 50 trials per variant. Inter-stimulus intervals were also randomized over the 50 trials per variant.

Procedure

At the beginning of the session participants were informed about the procedure and the different variants and performed some practice trials with all variants of the simulation. For each of the 50 trials per variant, participants first stated their prediction about the outcome of the run and attributed a certainty score to that prediction. They then watched the simulation and reported the temporal order they actually perceived. Participants started each trial themselves by pushing a start button so as to ensure that their attention was focused on the task. There were short breaks between variants. At the end of the session, participants were asked which modality, the visual or the auditory, they considered more accurate in perceiving minimal differences and whether they think that their prediction influenced their temporal order judgment. A testing session took between 60 and 90 minutes.

Data analyses

Means TOJ scores per variant and per ISI were compared with a repeated measures ANOVA. Paired-sample t-tests, using Bonferroni correction ($\alpha = .05/18 = .003$), were conducted to compare the TOJ scores for different ISIs across variants. A chi-square test of independence was performed to examine the relation between (strength of) prior belief and temporal order judgment. These analyses were conducted with the software packet PAWS Statistics 18.

In order to determine the threshold detection values of auditory and visual stimuli, psychometric functions were fitted to the data set of each variant using the MATLAB toolbox psignifit, version 2.5.6 (see http://bootstrap-software.org/psignifit/), which implements the maximum-likelihood method described by Wichmann and Hill (2001a). The threshold was defined as the point on the inter-stimulus interval axis corresponding to 75% correct responses, which is at the 50% point of the fitted psychometric function. Confidence intervals were obtained by the BCa bootstrap method implemented by psignifit, based on 1999 simulations (see Wichmann and Hill, 2001b).

Results

The differences between the mean TOJ scores were assessed with a repeated measure analysis of variance with variant and inter-stimulus interval as factors. Trials of the training session were not included in this analysis. Mauchly's test indicated that the assumption of sphericity had been violated for both main effects, variant $\chi^2(2) = 1.75$, $p < .05$ and inter-stimulus interval $\chi^2(14) = 64.39$, $p < .05$, and also for the variant x inter-stimulus interval interaction, $\chi^2(54) = 95.31$, $p < .05$. Therefore degrees of freedom were corrected using the Huynh-Feldt estimate of sphericity for the variant ($\varepsilon = 1.00$) and the Greenhouse-Geisser estimate of sphericity for the inter-stimulus interval ($\varepsilon = 0.47$) and variant x inter-stimulus interval interaction ($\varepsilon = 0.56$). Corrected mean scores per variant and inter-stimulus interval are shown in Table 1. The results of the two-way repeated measure ANOVA show that there was a significant main effect of variant, $F(2, 3.07) = 7{,}53$, $p = .001$, inter-stimulus interval, $F(2.37, 5.17) = 123.16$, $p < .001$, and a significant variant x inter-stimulus interval interaction, $F(5.57, 5.02) = 8{,}92$, $p < .001$. The results therefore suggest that the performance of subjects in correctly judging minimal differences were significantly different, depending on whether they received only visual or auditory input or both (see Figure 2).

Table 1
Corrected Mean Scores for Correct Temporal Order Judgments per Variant and per Inter-Stimulus Interval

	Inter-stimulus interval											
	20 ms		40 ms		60 ms		80 ms		100 ms		120 ms	
Variant	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual												
Correct judgment	.49	.26	.58	.17	.67	.20	.70	.16	.92	.15	.90	.18
Auditory												
Correct judgment	.22	.20	.31	.23	.55	.20	.70	.20	.91	.16	.98	.08
Visual & auditory												
Correct judgment	.40	.26	.49	.26	.58	.21	.72	.20	.94	.09	.96	.16

Note. The same 28 participants have taken part in all three variants.

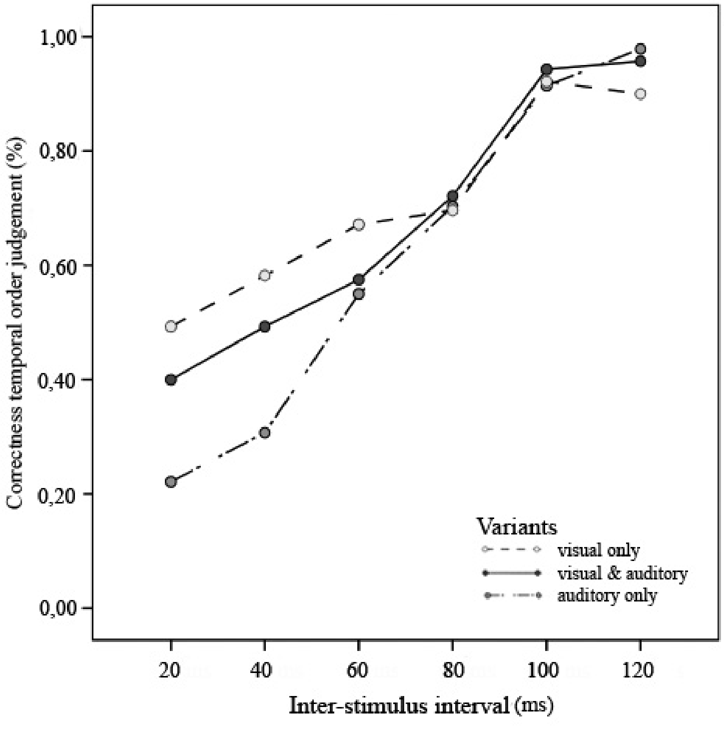

Figure 2. Corrected mean scores for correct temporal order judgments at different inter-stimulus intervals, depicted per variant.

Figure 2 shows that differences between variant were especially imminent in the lower inter-stimulus intervals. Post hoc t-tests revealed that the performance in the auditory variant was significantly lower than in the visual & auditory variant only at the 20 ms ISI (t (27) =

3.76, $p < .001$) and the 40 ms ISI (t (27) = 4.60, $p < .001$). The same holds for the auditory and visual variant where performance differences were significant at the 20 ms ISI (t (27) = 4.58, $p < 0.001$) and at the 40 ms ISI (t (27) = 6.60, $p < .001$). In all other cases, corrected mean scores did not differ significantly from each other (t (27) < 2.82, $p > .009$). Psychometric curves were fitted to the data of all three variants (see Figure 3). The temporal order thresholds of all three variants lie pretty close together. TOT values ranged from 73.28 ms (95% BC_a [69.32 ms, 71.30 ms]) for the visual variant, through 79.76 ms (95% BC_a [77.86 ms, 81.61 ms]) for the visual & auditory variant, to 83.44 ms (95% BC_a [81.47ms, 86.67 ms]) for the auditory only variant. TOJs for values that fall under this TOT cannot be correctly identified above chance level.

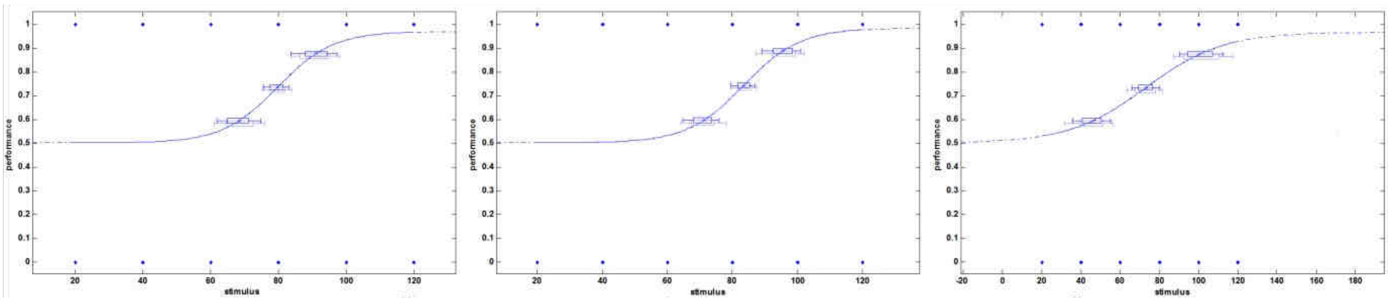

Figure 3. Psychometric curves depicted with temporal order threshold at 50% above guessing level, with 20% and 80% cutoff and BCa error bars. From left to right: visual & auditory variant, auditory only variant, visual only variant.

When considering the overall differences between variants, 30 percent of the participants reported after the experiment that they thought they performed best in the visual modality, whereas 63 percent thought so of the auditory modality. For 6 percent there was no subjective difference between both modalities.

The number of correct temporal order judgments as a function of correct prior beliefs is shown in the left-hand side of Figure 4 and the number of incorrect TOJs as a function of incorrect prior beliefs in the right-hand side. Participants with correct prior beliefs were more likely to make a correct temporal order judgment than with an incorrect prior belief, whereas a incorrect prior belief did not matter. The Chi-square test for Cross Tabulation confirmed that the differences were highly significant, $\chi^2(1, N = 4200) = 86.317$, $p < 0.001$, $\varphi = 0.14$. Participants were also aware of this influence, however not very strongly, $ø = 2.96$ and $SD = 0.64$ (on a scale from 1 = 'not at all' to 4 = 'a lot').

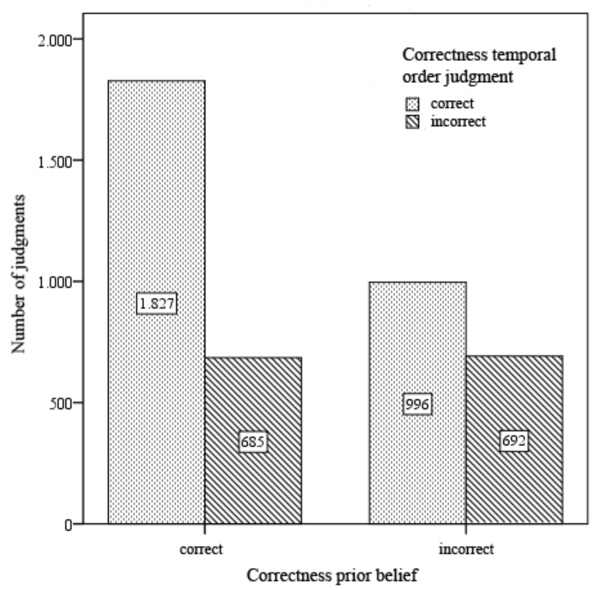

Figure 4. Number of correct and incorrect temporal order judgments based on the correctness of prior beliefs

Also, the strength of the prior belief seems to matter. When the chi-square test was conducted on subgroups created by certainty rating, the prior belief and temporal order judgment variables did not always depend on each other (see Figure 5). The tendency to make more correct temporal order judgments when the prior belief was correct was only observable when the participant rated their certainty with 3 ($\chi^2(1, N = 1643) = 27.334$, $p < 0.001$, $\varphi = 0.13$) or 4 ($\chi^2(1, N = 1396) = 53.032$, $p < 0.001$, $\varphi = 0.20$). Due to the small number of ratings with a certainty of 5, the chi-square test was significant only at alpha level of $\alpha = 0.010$ ($\chi^2(1, N = 190) = 6.692$, $p = 0.010$, $\varphi = 0.19$).

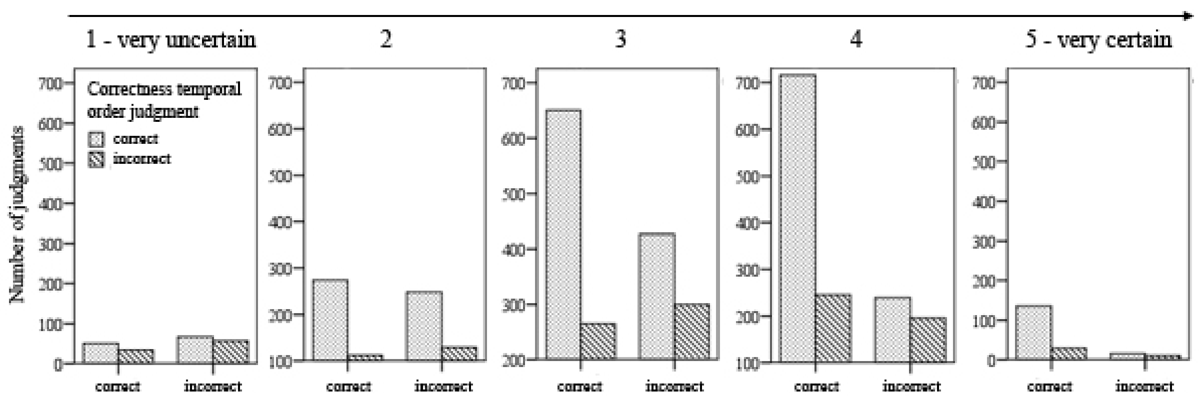

Figure 5. Number of correct and incorrect temporal order judgments based on the correctness of prior beliefs per certainty subgroup.

Conclusion and discussion

Goal of this research was to explore how the process of observing evidence in inquiry learning tasks takes places and what can impede learners in making good observations. It was therefore examined whether the perception of minimal differences in evidence observation differs between the visual and auditory modality and what the influence of cross-modal interaction and (the strength of) prior beliefs is.

Identifying minimal temporal differences

The results of the present research show that, with ambiguous stimuli, the human information processing system is sometimes unable to detect minimal differences. It was expected that typical results of TOJ task could be replicated in the present study. This was in fact the case. In accordance with general assumptions from psychophysical research, there is no absolute threshold where differences suddenly can be detected correctly. Minimal temporal differences can only be identified very occasionally whereas very distinct differences almost always can. For all three variants, visual only, auditory only and visual & auditory, psychometric functions can be established that graphically correspond with these assumptions.

Temporal order thresholds values of about 80 ms for both, the auditory and the visual & auditory variant, are found. In the visual variant, participants scored slightly better. It can be concluded that very small visual and auditory differences that fall under these thresholds cannot be detected above chance level in either variant. But despite the better performance in the visual variant, overall threshold values are still rather high, especially compared to less complex sounds like for example TOJs with two auditory stimuli of different tone frequencies. Here, thresholds values of as low as 10 ms were found (Fink et al., 2005). Although stimuli characteristics in inquiry learning tasks will vary widely, it is likely that similar results will also hold true for other inquiry tasks. Stimuli in inquiry learning tasks are in general moderately to highly complex.

Differences between TOJs of auditory and visual stimuli

It was expected that auditory stimuli are superior in temporal processing and therefore better performance in TOJ tasks can be anticipated in the auditory variant. However, the contrary is true. Participants performed better in the visual variant, both in terms of threshold values and of detecting very small differences. Strikingly, even the participants do not expect this and report after the experiment that they think they made more accurate judgments in the auditory variant.

An explanation for the superiority of the visual only presentation might be found in the inquiry-specific context of the temporal order judgments. In typical TOJ tasks, there is always a point at which the visual stimuli suddenly appear. In inquiry learning task visual stimuli move continuously along a certain path before the temporal order event happens, as is the case in the rock-dropping experiment Chinn and Malhotra (2002). This adds a spatial dimension to the temporal order judgment of visual stimuli. The moving cars are always visible, which makes it possible for the participants to partly estimate the temporal order in which the cars will crash in advance. It is well known that the visual system uses the predictability of the path of a moving object to calculate positions in advance and thereby correct for the visual latency (Krekelberg & Lappe, 2001).

Cross-modal interaction effects

Cross-modal interaction effects are common in psychophysical tasks and it was expected that a positive distortion of TOJ performance scores at the presence of both visual and auditory stimuli was likely to occur. However, as the results of the present study indicate, that having access to both visual and auditory stimuli for making judgments does not strengthen the accuracy of temporal order judgments. Participants even perform slightly worse in the visual & auditory variant, compared to the visual only variant, albeit not to a statistically significant degree. It can therefore be inferred that no cross-modal interaction effects are detected in this study. Participants in the auditory & visual variant do not score any better (or worse) than in the best variant, visual only.

Influence of prior beliefs

It was expected beforehand that participants' prior beliefs would influence the subsequent temporal order judgment. This is in fact what the results of the present study indicate. The asymmetrical bias that Chinn and Malhotra (2002) found in their rock-dropping experiment is replicated in this research. Participants with correct prior beliefs are more likely to make correct temporal order judgments, whereas participants with wrong prior beliefs neither make more or less correct judgments. Before conducting this study, it was not sure whether and how the strength of those prior beliefs influences the subjective experience of minimal differences. The results of the present study show that the influence of correct prior beliefs is even stronger when participants are more certain about their prediction. In the experiment at hand, prediction and temporal order judgment directly succeed each other. It is therefore not surprising that participants are also aware of the influence of their predictions. But as the study of Fernbach and colleagues (2007) has shown, even unrelated events might influence the observation of evidence, without learners being aware of it. A plausible explanation for the influence of bias on the TOJs might be that participants focus more on the car which they think will crash first. They can therefore make more accurate judgements when their prediction and the real temporal order match.

Practical implications

Having a better understanding of what can go wrong during the process of observing evidence in inquiry learning can help to develop support for learners. The main practical implication of this study is that learners should be made aware of their limited capacity and be supported in making better observations. When differences are smaller than 80 ms, multiple observations should be made and inquiry learners should not focus on the auditory stimuli but rather focus on visual stimuli only or on both auditory and visual stimuli. This might explain why attempts to promote evidence evaluation in the study of Chinn & Malhotra (2002) by telling the children to 'listen carefully' did not have any positive effects. No special attention has to be granted for supporting learners in making observations at relatively large ISIs above 80 ms.

It seems that being aware of one's prior beliefs can lead to better observations in many cases. Inquiry learners should therefore be forced to express their ideas about the outcomes of an experiment beforehand. In case that they have correct prior beliefs, this will lead to better evidence observation. Making an incorrect prediction would yet do no harm. To take this argument even further, this implies that explorative experimentation cannot be very effective in the context of inquiry learning. The same conclusion has already been drawn in other research contexts (see for example Adam, 2002; Lazonder, Wilhelm & Hagemans, 2008).

One should however not reason that it is better to provide learners with ready-made results. On the one hand, it is a false conclusion to think that merely reading off results cannot be impeded. Research has shown that even directly reading of empirical results of measurement and experimentation can be confounded by prior beliefs (Adam, 2002). And on the other hand, for effective inquiry learning, it is important to let learners directly observe evidence in contrast to second-hand evidence (Zimmerman, 2007). The ability to make good observations lies at the heart of learning and doing science. Or to put it into the words of George Santayana (1863 - 1952), one of the most influential American philosophers of the

20th century: “Science is nothing but developed perception, interpreted intent, common sense rounded out and minutely articulated.”

References

Adam, M. (2002). *Theoriebeladenheit und Objektivität. Zur Rolle von Beobachtungen in den Naturwissenschaften* [Theory-ladenness and objectivity. On the role of observations in science], Frankfurt am Main: Hänsel-Hohenhausen.

Aschersleben, G., & Müsseler, J. (1999). Dissociations in the timing of stationary and moving stimuli. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(6), 1709-1720.

Bergwanger, D., Wittmann, M., von Steinbüchel, N. & von Suchodoletz, W. (2004). Measurement of temporal-order judgment in children. *Acta Neurobiologiae Experimentalism, 64*(3), 387-394.

Binda, P., Morrone, M. C., & Burr, D. C. (2010). Temporal auditory capture does not affect the time course of saccadic mislocalization of visual stimuli. *Journal of Vision*, *10*, 1-13.

Chinn, C., & Malhotra, B. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, *94*(2), 327–343.

Dai, H. (1995). On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods. *The Journal of the Acoustical Society of America, 98*(6), 3135-3139.

Durant, J. (1994). What is scientific literacy? *European Review, 2*(1), 83-89.

Fernbach, P. M., Linson-Gentry, P., & Sloman, S. A. (2007). Causal Beliefs Influence the Perception of Temporal Order. In *Proceedings of the twenty-ninth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Earlbaum.

Fink, M., Churan., J., & Wittmann, M. (2005). Assessment of auditory temporal-order thresholds – a comparison of different measurement procedures and the influences of age and gender. *Restorative Neurology and Neuroscience, 23,* 281-296.

Florentine, M., Buus, S., & Geng, W. (1999). Psychometric functions for gap detection in a yes-no procedure. *The Journal of the Acoustical Society of America*, *106*(6), 3512-20.

Formby, C., Gerber, M. J., Sherlock, L. P., & Magder, L. S. (1998). Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection. *The Journal of the Acoustical Society of America*, *103*(6), 3554-60.

Irwin, R., Ball, A., Kay, N., Stillman, J., & Rosser, J. (1985). The development of auditory temporal acuity in children. *Child Development*, *56*(3), 614–620.

Jaśkowski, P., Jaroszyk, F., & Hojan-Jezierska, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research*, *52*, 35 -38.

Klahr, D. & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science, 12*, 1-48.

Klahr, D., Fay, A., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*, 111-146.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674-689.

Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics*, *63*(8), 1377-1388.

Klein, S. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, *63*(8), 1421-51.

Krekelberg, B., & Lappe, M. (2001). Neuronal latencies and the position of moving objects. *Trends in Neurosciences*, *24*(6), 335–339.

Lazonder, A.W., Wilhelm, P. & Hagemans, M.G. (2008). The influence of domain knowledge on strategy use during simulation-based inquiry learning. *Learning and Instruction, 18* (6), 580-592.

Moore, B. C. (1982). *An introduction to the psychology of hearing* (2 ed.). London: Academic Press.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, *1*(2), 56–61.

Sternberg, S. & Knoll, R.L. (1973). The perception of temporal order: Fundamental issues and a general model. In: S. Kornblum (eds), *Attention and performance* (Vol. 4) (pp. 629–685), Academic Press: New York.

Szelag, E., Kanabus, M., Kolodziejczyk, I., Kowalska, J., & Szuchnik, J. (2004). Individual differences in temporal information processing in humans. *Acta Neurobiologiae Experimentalis*, *64*(3), 349-66.

Ulbrich, P., Churan, J., Fink, M., & Wittmann, M. (2009). Perception of temporal order: the effects of age, sex, and cognitive factors. *Aging, Neuropsychology, and Cognition*, *16*(2), 183-202.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research, 22*(1), 32-35.

Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio–visual integration in temporal perception. *International Journal of Psychophysiology, 50*(1-2), 117-124.

Werner, L. (1996). The development of auditory behavior (or what the anatomist and physiologists have to explain). *Ear & Hearing*, *17*(5), 438-446.

Wichmann, F.A., & Hill, N.J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*, *63*(8), 1293–1313.

Wichmann, F.A., & Hill, N.J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314-29.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental* Review, 27, 172-223.