

# MODELING AND EXPERIMENTAL STUDY OF CONFIDENCE IN SENSORY JUDGMENTS

Valentin M. Shendyapin, Irina G. Skotnikova

*Moscow State University of Psychology and Education, Institute of Psychology Russian Academy of Sciences, Moscow, Russia*

*valshend@yandex.ru, iris236@yandex.ru*

## Abstract

*A mathematical model of decision making and confidence in sensory tasks is developed on the base of Signal Detection Theory. The model argues that confidence may be represented as a sum of a sensory evidence and non-sensory (frequency and motivational) ones (determined by stimuli a priori probabilities and costs—payoffs for responses). This representation of confidence permits to convert confidence obtained in a single observation into expected decision correctness probability and utility of a chosen response. The model predicts better accuracy of confident responses regarding to all responses. It was confirmed in the both main kinds of sensory discrimination tasks: “Greater—Lesser” and “Same—Different” for visual stimuli: spatial (circles diameters) and temporal (durations) ones presented simultaneously and successively correspondingly. Thus, the fact found seems to be rather general one.*

A number of mathematical models describe decision making and confidence in sensory discrimination. They are referred as developed in the framework of Signal Detection Theory (SDT), Sequential—Sampling models, Accumulator one and Doubt Scaling one (see Baranski, Petrusic, 1998 for the review). And an additional class of models may be selected — Neural Network ones (see Shendyapin, Skotnikova, 2001 for the review). These models suggest productive approaches for description of confidence internal feeling. However they do not include theoretical formulas to convert an internal representation of decision confidence (available for an observer only) into external one expressed by means of decision correctness probability in a single observation, which may be registered by an experimenter. At the same time an observer makes this conversion easily. So, it is very desirable to explain how he (she) can carry out the conversion mentioned.

The aim of the study presented is to find a suitable mathematical representation of confidence in the framework of SDT (Green, Swets, 1974) which would permit to express a correctness probability of a hypothesis chosen and the corresponding response utility as a function of confidence.

## Decision making and confidence in the task of the most probable hypothesis choice

An a posteriori probability of signal occurrence in a single observation is given by  $P(\mathbf{sn}|x) = [l(x) l_0]/[1 + l(x) l_0]$  (Egan, 1975), where:  $\mathbf{sn}$  — a signal presentation,  $x$  — a sensory effect value,  $l(x) = f(x|\mathbf{sn})/f(x|\mathbf{n})$  — likelihood ratio,  $l_0 = P(\mathbf{sn})/P(\mathbf{n})$  — ratio of signal to noise a priori probabilities,  $\mathbf{n}$  — a noise presentation. After mathematical transformation it becomes  $P[\mathbf{sn}|C(x)] = 0.5 + 0.5 \operatorname{th}[C(x)/2]$  where  $C(x) = L_0 + L(x)$ ,  $L_0 = \ln(l_0)$ ,  $L(x) = \ln[l(x)]$  (Shendyapin et al., 2010). Figure 1 shows that a correctness probability of  $H_s$ - hypothesis (“The signal has been presented”) increases monotonically from 0 to 1 while  $P[\mathbf{n}|C(x)] = 0.5 - 0.5 \operatorname{th}[C(x)/2]$  — a correctness probability of  $H_n$ -hypothesis (“The noise has been

presented”) decreases monotonically from 1 to 0 according to variable  $C$  increase (sensory effect  $x$  will be omitted for convenience further).

A decision making rule in the task of a most probable hypothesis choice is: if  $C > 0$  then  $H_s$ -hypothesis has to be chosen because  $P(\mathbf{sn}, \mathbf{Y}|x) = 0.5 + 0.5 \text{th}[C(x)/2]$  — correctness probability of the response  $\mathbf{Y}$  (“Yes, the signal has been presented”) is more than 0.5, if  $C < 0$  then  $H_n$ -hypothesis has to be chosen because  $P(\mathbf{n}, \mathbf{N}|x) = 0.5 - 0.5 \text{th}[C(x)/2]$  — correctness probability of the response  $\mathbf{N}$  (“No, the noise has been presented”) is more than 0.5.

Thus positive value of the variable  $C$  looks as an evidence in favor of  $H_s$ -hypothesis correctness while its negative value — as an evidence in favor of  $H_n$ -hypothesis correctness.  $C$  is a sum of 2 items:  $L_0$  and  $L(x)$ . Therefore positive value of  $L_0 = \ln[P(\mathbf{sn})/P(\mathbf{n})]$  may be considered as a *non-sensory frequency evidence* (determined by a priori probability of signal and that of noise) in favor of  $H_s$ -hypothesis correctness while negative one — as appropriate evidence in favor of  $H_n$ -hypothesis correctness. As well, positive value of  $L(x) = \ln[f(x|\mathbf{sn})/f(x|\mathbf{n})]$  (determined by a posteriori likelihood ratio obtained in a current observation) may be considered as a *sensory evidence* in favor of  $H_s$ -hypothesis correctness while negative one — as appropriate evidence in favor of  $H_n$ -hypothesis correctness.

A decision making criterion for studied task is  $C_{cr} = 0$ . An appropriate representation of confidence  $Con_{Ycor}$  in  $\mathbf{Y}$ -response correctness may be taken as a distance from an obtained point  $C$  (which has positive value in this case) to the criterion  $C_{cr}$  in the  $C$ -axis. So,  $Con_{Ycor} = C$ . And confidence  $Con_{Ncor}$  in  $\mathbf{N}$ -response correctness may be taken as a distance from the criterion  $C_{cr}$  to the point  $C$  (which has negative value in this case). So,  $Con_{Ncor} = -C$ .

Suggested representations of confidences  $Con_{Ycor}$  and  $Con_{Ncor}$  are based on a sum of the both frequency evidence and sensory one. It corresponds to confidence representations in Sequential—Sampling models (Heath, 1984; Link, 1992), Accumulator one (Vickers et al., 1998, 2003), Doubt Scaling one (Baranski, Petrusic, 1998). Treating of confidence value as measured by a distance from an obtained point  $C$  to a criterion  $C_{cr}$  corresponds to understanding of confidence value as a distance between a criterion and a current sensory effect, which is developed in the frameworks of the previous SDT-models of confidence (Balakrishnan, Ratcliff, 1996; Bjorkman et al., 1993). Main features of the model proposed in the present study are: *the content of evidences was introduced into SDT paradigm, non-sensory evidence was taken into account in addition to sensory one and a positive relation was proved mathematically between confidence and probability of correct responses in a single observation.*

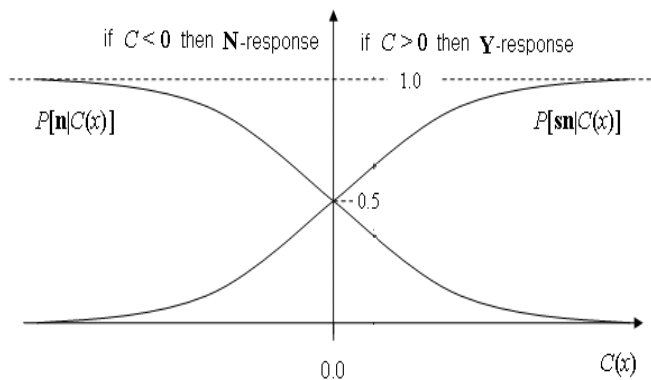


Figure 1. Theoretical probabilities of  $H_s$ - and  $H_n$ -hypothesis correctness as functions of  $C$  which is the sum of the frequency evidence  $L_0$  and the sensory one  $L(x)$ .

The theoretical functions derived (see Figure 1) correspond to a sigmoid function obtained empirically which describes a relation between confidence category and proportion correct (see Bjorkman et al, 1993; Link, 1992 for reviews). The correspondence between theoretical and empirical functions confirms adequacy of the theoretical ones.

## Decision making and confidence in the task of the most useful response choice

The model suggested is able to represent confidence in case when each response choice brings the appropriate useful result  $V$  to an observer. The discrete values of the result  $V$  depend on correctness/erroneousness of this response. The results of the correct responses ( $v_{sn,Y}$ ,  $v_{n,N}$ ) are positive (useful) while those of erroneous ones ( $v_{sn,N}$ ,  $v_{n,Y}$ ) are negative (wasteful). In each observation an observer has to choose a response giving him the greatest utility.

It was shown that *utility (defined as the average value of the result  $V$ )* caused by  $\mathbf{Y}$ -response is  $E[V(\mathbf{Y}|C)] = 0.5 (v_{sn,Y} + v_{n,Y}) + 0.5 (v_{sn,Y} - v_{n,Y}) th(C/2)$  and utility caused by  $\mathbf{N}$ -response is  $E[V(\mathbf{N}|C)] = 0.5 (v_{sn,N} + v_{n,N}) + 0.5 (v_{sn,N} - v_{n,N}) th(C/2)$  (Shendyapin et al., 2010). Figure 2 shows two examples of these utilities as functions of variable  $C$  for 2 sets of the result  $V$  discrete values. Utility of  $\mathbf{Y}$ -response increases monotonically from negative  $v_{sn,N}$  to positive  $v_{sn,Y}$  while utility of  $\mathbf{N}$ -response decreases monotonically from positive  $v_{n,N}$  to negative  $v_{sn,N}$  according to the variable  $C$  increase. The curves representing the utilities intersect at the point  $C_{cr} = -L_V$ , where  $L_V = \ln[(v_{sn,Y} - v_{sn,N})/(v_{n,N} - v_{n,Y})]$  depends on values of the result  $V$ . This point serves as a decision making criterion for a task mentioned. The utility obtained in single observation reaches its minimum  $E(V)_{min}$  in this point.

A *decision making rule* in the task of the most useful response choice is: if a current sensory effect  $x$  gives inequality  $C > -L_V$  then  $\mathbf{Y}$ -response has to be chosen because its utility  $E[V(\mathbf{Y}|C)]$  is the greatest, if  $C < -L_V$  then  $\mathbf{N}$ -response has to be chosen because its utility  $E[V(\mathbf{N}|C)]$  is the greatest (see Figure 2).

Inequality  $C + L_V > 0$  follows from inequality  $C > -L_V$  f, where  $C = L_0 + L(x)$ . So, a decision making rule may be reformed: a positive value of the sum  $L_0 + L(x) + L_V$  may be considered as an evidence in favor of  $\mathbf{Y}$ -response choice while a negative one — as an evidence in favor of  $\mathbf{N}$ -response choice. Let's look at the case when an observer obtains a current sensory effect  $x$  which makes the sum  $L_0 + L(x)$  equal to zero. In this case the positive value of the rest item  $L_V$  may be considered as evidence in favor of  $\mathbf{Y}$ -response choice, while negative one — as evidence in favor of  $\mathbf{N}$ -response choice. Recalling that positive value of  $L_V$  means  $v_{sn,Y} - v_{sn,N} > v_{n,N} - v_{n,Y}$  we may conclude that the observer chooses  $\mathbf{Y}$ -response because he attaches importance rather to signal than to noise. Therefore the third item  $L_V$  represents *a special kind of non-sensory evidence* (determined by costs—payoffs) in favor of either  $\mathbf{Y}$ - or  $\mathbf{N}$ -response choice. It may be called as *motivational evidence* which represents an observer's motivation to choose a response corresponding to a more significant stimulus, in distinction of frequency evidence and sensory one which represent decision correctness probability.

Confidence  $Con_{Yutil}$  in the greatest utility of  $\mathbf{Y}$ -response may be represented as a distance from  $C$ -value obtained in a current observation to the criterion  $C_{cr} = -L_V$ . Thus  $Con_{Yutil}$  equals to the sum of the frequency evidence, sensory one and motivational one:  $L_0 + L(x) + L_V$ , because this sum is positive for  $\mathbf{Y}$ -response choice. Confidence  $Con_{Nutil}$  in the greatest utility of  $\mathbf{N}$ -response is defined as a distance from  $C_{cr} = -L_V$  to  $C$ . Thus  $Con_{Nutil}$  equals to  $-(L_0 + L(x) + L_V)$  because the sum of evidences  $L_0 + L(x) + L_V$  is negative for  $\mathbf{N}$ -response choice.

## Decision making and confidence in the task of the greatest and positive utility choice

The model suggested permits to represent confidence in a more elaborate task, when a response chosen has to bring utility which is the greatest and positive at the same time. In the case when an observer feels that the utilities of the both  $\mathbf{Y}$ - and  $\mathbf{N}$ -response are negative and he is in doubt how to proceed then he has to choose the cautious  $\mathbf{D}$ -response (“Doubt”).

There are two special points in the  $C$ -axis, which are related with the positive sign of responses utility. Monotonically increasing utility  $V(\mathbf{Y}|C)$  equals to 0 in the first point  $C_{crY}$ . Therefore we have equation  $V(\mathbf{Y}|C_{crY}) = 0$ , from which  $C_{crY} = -L_{VY}$  can be derived where  $L_{VY} = \ln(-v_{sn,Y}/v_{n,Y})$ . This point serves as a criterion for the utility sign of  $\mathbf{Y}$ -response: if  $C > -L_{VY}$  then the utility of  $\mathbf{Y}$ -response has positive sign, if  $C < -L_{VY}$  then it has negative sign. Monotonically decreasing utility  $V(\mathbf{N}|C)$  equals to 0 in the second point  $C_{crN}$ , therefore we have equation  $V(\mathbf{N}|C_{crN}) = 0$  and its solution  $C_{crN} = -L_{VN}$ , where  $L_{VN} = \ln(-v_{sn,N}/v_{n,N})$ . This point serves as a criterion for the utility sign of  $\mathbf{N}$ -response: if  $C < -L_{VN}$  then the utility of  $\mathbf{N}$ -response has positive sign, if  $C > -L_{VN}$  then it has negative sign.

Criterion  $C_{cr}$  is always between criteria  $C_{crY}$  and  $C_{crN}$ . But we have two cases in dependence of given values  $v_{sn,Y}$ ,  $v_{sn,N}$ ,  $v_{n,N}$ ,  $v_{n,Y}$ . Even the utility minimum  $E(V)_{min}$  is positive in the case of low costs, i.e.  $v_{n,Y} v_{sn,N}/v_{sn,Y} v_{n,N} < 1$ , when  $C_{crN} > C_{crY}$  (see Figure 2, a). In this case an observer always may be confident because the utilities of the both  $\mathbf{Y}$ - and  $\mathbf{N}$ -response chosen are positive in the whole  $C$ -axis. A *decision making rule* in the task of the greatest and positive utility choice for this case is: if  $C > C_{cr}$  then confident  $\mathbf{Y}$ -response has to be chosen, if  $C < C_{cr}$  then confident  $\mathbf{N}$ -response has to be chosen (see Figure 2, a). Confidences  $Con_{Yutil+}$  and  $Con_{Nutil+}$  in the greatest and positive utility of  $\mathbf{Y}$ - and  $\mathbf{N}$ -response are equal to above  $Con_{Yutil}$  and  $Con_{Nutil}$  as a matter of fact.

The utility minimum  $E(V)_{min}$  is negative and in the neighborhood of the criterion  $C_{cr}$  a caution interval  $(C_{crN}, C_{crY})$  arises where the utilities of the both  $\mathbf{Y}$ - and  $\mathbf{N}$ -response chosen are negative, in the case of high costs, i.e.  $v_{n,Y} v_{sn,N}/v_{sn,Y} v_{n,N} > 1$ , when  $C_{crN} < C_{crY}$ , (see Figure 2, b). Therefore a *decision making rule* in the task of the greatest and positive utility choice for this case is: if  $C > C_{crY}$  then confident  $\mathbf{Y}$ -response has to be chosen, if  $C < C_{crN}$  then confident  $\mathbf{N}$ -response has to be chosen, if  $C_{crN} < C < C_{crY}$  then cautious  $\mathbf{D}$ -response has to be chosen (see Figure 2, b).

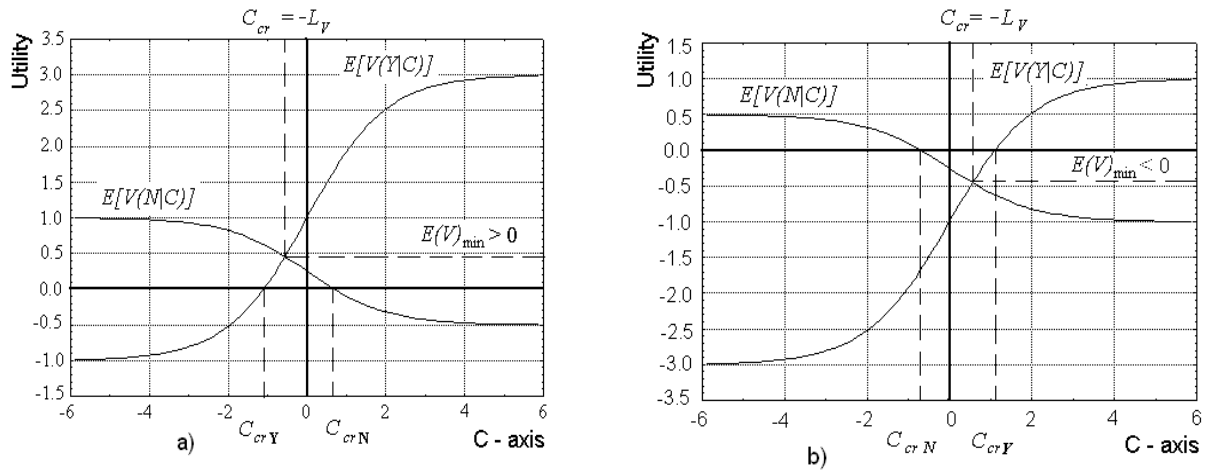


Figure 2. The utilities  $E[V(\mathbf{Y}|C)]$  and  $E[V(\mathbf{N}|C)]$  as functions of the variable  $C$ .

a)  $v_{sn,Y} = 3$ ,  $v_{n,Y} = -1$ ,  $v_{n,N} = 1$ ,  $v_{sn,N} = -0.5$ ,  $E(V)_{min} = +0.4545$ ,  $C_{cr} = -\ln(3.5/2)$ ,  $C_{crN} = \ln(2)$ ,  $C_{crY} = -\ln(3)$ .

b)  $v_{sn,Y} = 1$ ,  $v_{n,Y} = -3$ ,  $v_{n,N} = 0.5$ ,  $v_{sn,N} = -1$ ,  $E(V)_{min} = -0.4545$ ,  $C_{cr} = \ln(3.5/2)$ ,  $C_{crN} = -\ln(2)$ ,  $C_{crY} = \ln(3)$ .

Confidence  $Con_{Yutil+}$  in the greatest and positive utility of  $\mathbf{Y}$ -response for this case may be represented as a distance from  $C$ -value obtained in a current observation to the criterion  $C_{crY} = -L_{VY}$ . Thus  $Con_{Yutil+}$  equals to the sum of the frequency evidence, sensory one and item  $L_{VY}$ :  $L_0 + L(x) + L_{VY}$ , because this sum is positive for  $\mathbf{Y}$ -response choice.

Confidence  $Con_{Nutil+}$  in the greatest and positive utility of N-response is defined as a distance from  $C_{crN} = -L_{VN}$  to  $C$ . Thus  $Con_{Nutil+}$  equals to  $-(L_0 + L(x) + L_{VN})$  because the sum  $L_0 + L(x) + L_{VN}$  is negative for N-response choice.

### **An experimental examination of the model prediction**

The model described predicts higher proportion of correct responses among confident ones regarding to overall proportion correct. This prediction was examined empirically using data of visual discrimination: a) of 2 circles diameters simultaneously presented, 93 experiments of 200–800 trials each, in 49 participants, “Greater—Lesser” (“>,<”) task including “Doubt”-responses (data of Shendyapin et al, 2010); b) of 2 light flashes durations successively presented, 98 experiments of 100 trials each, in 71 participants, “Same—Different” (“=,≠”) task including 2-categories confidence evaluation (“Confident—Unconfident”) as a 2nd response (summary data of Skotnikova, 1994; Golovina, Skotnikova, 2010). The instruction used in “>,<”-task paid attention to losses arisen from erroneous responses while the instruction used in “=,≠”-task did not. In each task a difference between stimuli compared was chosen around an individual’s difference threshold level, therefore PC-values (overall proportions of correct responses) were obtained as 0.54—0.8 and 0.7—0.8 correspondingly.

As a result, higher proportions of correct responses among confident ones as compared to overall proportions correct were obtained in 87% of experiments using “>,<”-task and in 85% of experiments using “=,≠”-task ( $p < 0.01$  in the both tasks, Sign Test). Twice better accuracy improvement ( $p < 0.0003$ , Mann—Whitney Test) was observed in participants who were warned about losses arisen from erroneous responses in comparison with those who were not warned. Thus the greatest accuracy improvement seems caused by the instruction which stimulated participants to reach high level of confidence of decisions correctness. Advantage found points out to efficiency of a conscious regulation of a sensory task performing with the help of confidence.

Exploratory experiment has showed that the value of the accuracy improvement correlated positively ( $p < 0.039$ ) with reflection — impulsivity index estimated via Matching Familiar Figures Test (Kagan, 1966). Factor analysis conducted has confirmed this result which revealed better accuracy improvement in reflective persons in comparison with impulsive ones.

Moreover, data of another authors were selected which were obtained in sensory discrimination tasks including 2 categories of confidence evaluation (“Confident—Unconfident”) which allowed to analyze them in the form analogous to our data analysis: discrimination of visual depth and of sound localization, “Yes—No” design,  $PC = 0.75$  (Obrink, 1948); discrimination of visual speed, 2 Alternative Forced Choice design,  $PC = 0.783—0.793$  (Bjorkman, Qvarsell, 1963). Again greater proportions of correct responses among confident ones were obtained regarding to overall proportions correct. It confirms the fact revealed in the present authors experiments and appeared to be rather general one, and confirms the prediction of the model developing as well. It means that a majority of an observer’s errors “passes“ into a sample of unconfident responses and therefore the rest sample of confident ones contains more correct responses than the whole data sample does. A clear picture of a typical unconfidence of erroneous responses was given by our study mentioned above: proportions of unconfident responses among erroneous ones were greater than those among correct ones (Skotnikova, 1994). All the data shown correspond to the classical fact of proportion correct increase along with confidence category increase (see Bjorkman et al, 1993 for a review and for original data) and elucidate different aspects of confidence—accuracy relation.

## References

- Golovina, E.V., Skotnikova, I.G. (2010). Kognitivno–stilevaya structura uverennosti. *Psychologicheskii Journal*, t. 1, № 4, s. 69–82. In Russian. (Structure of confidence dependent on cognitive styles. *Psychological Journal*, 31 (4), 69–82).
- Balakrishnan, J.D., Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception & Performance*, 22, 615–633.
- Baranski, J.V., Petrusic, W.M. (1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception & Performance*, 24, 929–945.
- Bjorkman, M., Juslin, P., Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics*, 54, 75–81.
- Bjorkman, M., Qvarsell, B. (1963). Confidence in S—E predictions. *Reports from the Psychological Laboratory, University of Stockholm*, University of Stockholm: Stockholm, № 155.
- Egan, J.P. (1975). Signal Detection Theory and Receiver Operation Characteristic. N.Y.: Academic Press.
- Ferrel, W.R. (1995). A model for realism of confidence judgments: implications of underconfidence in sensory discrimination. *Perception & Psychophysics*, 57, 246–254.
- Green, D.M., Swets, J.A. (1974). Signal Detection Theory and Psychophysics. N.Y.: Wiley, 2nd ed.
- Kagan, J. (1966). Reflection—impulsivity: the generality of dynamics of conceptual tempo. *Journal of Abnormal Psychology*, 71, 17–24.
- Heath, R.A. (1984). Random—walk and accumulator models of psychophysical discrimination: a critical evaluation. *Perception*, 13, 57–65.
- Link, S.W. (1992). The wave theory of difference and similarity. Hillsdale, N.Y.: Erlbaum.
- Obrink, J. (1948). An experimental investigation of confidence. Ph. D. Thesis, Uppsala University: Uppsala.
- Shendyapin, V.M., Barabanshikov, V.A., Skotnikova, I.G. (2010). Uverennost v reshenii: modelirovanie i experimentalnaya proverka. *Experimentalnaya psichologiy*, t. 3. № 1. S. 30–57. In Russian. (Confidence in decision making: modeling and experimental examination. *Experimental Psychology*, 3 (1), 30–57).
- Shendyapin, V.M., Skotnikova, I.G. (2001). Neural network model as a possible tool for modeling of confidence in sensory judgments. *Fechner Day 2001: Proceedings of the 17th Annual Meeting of the International Society for Psychophysics*, Leipzig: PABST, 597–601.
- Skotnikova, I.G. (1994). I. Accuracy, confidence—unconfidence and response times in unordered discrimination: subject-oriented research. II. Confidence judgments specificity in same—different procedure (visual duration discrimination). *Fechner Day 1994: Proceedings of the 10th Annual Meeting of the International Society for Psychophysics*, University of British Columbia: Vancouver, 208–219.
- Vickers, D. (2003). Confidence and response time in three-category judgment. *Fechner Day 2003: Proceedings of the 13th Annual Meeting of the International Society for Psychophysics*, Jannes sanbbtryck: Stockholm, 325–330.
- Vickers, D., Lee, M. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator model. *Nonlinear dynamics, psychology and life sciences*, 2, 169–194.