It's a Catastrophe! Testing dynamics between competing cognitive states using mixture and hidden Markov models

Visser, I.; Speekenbrink, M.

Publication date
2014

Document Version
Final published version

Published in
COGSCI 2014

It's a Catastrophe! Testing dynamics between competing cognitive states using mixture and hidden Markov models

Ingmar Visser (i.visser@uva.nl)
Department of Developmental Psychology, University of Amsterdam
Weesperplein 4, 1018 XA Amsterdam, The Netherlands

Maarten Speekenbrink (m.speekenbrink@ucl.ac.uk)
Department of Experimental Psychology, University College London
Gower Street, London WC1E 6BT, United Kingdom

Abstract

Dual or multiple systems approaches are ubiquitous in cognitive science, with examples in memory, perception, categorization, cognitive development, and many other fields. Dynamical systems models with multiple stable states or modes of behavior are also increasingly used in explaining cognitive phenomena. Catastrophe theory provides a formal framework for studying the dynamics of switching between two qualitatively distinct modes of behavior. Here we present a parametric approach to testing specific predictions about the dynamics of such switches that follow from catastrophe theory. These so-called catastrophe flags are bimodality, divergence, and hysteresis. We show how these three flags can be tested using (constrained) mixture and hidden Markov models and provide an example of each using three different data sets.

Keywords: hidden Markov model; mixture model; dynamical systems; catastrophe theory; stagewise development; bimodality; divergence; hysteresis.

Introduction: competing states

In cognitive science, theories of learning, change, and development often involve multiple systems or multiple stages. During category learning, participants may switch from rule-based to exemplar-based learning (Johansen & Palmeri, 2002), or switch between alternative hypotheses about the category structure when learning through hypothesis testing. In social psychology, people adapt their attitudes when they become more involved (van der Maas et al., 2003), sometimes leading to polarization. Piagetian developmental theory predicts that children develop through a number of cognitive stages, which are characterized by specific behavioral strategies (Inhelder & Piaget, 1958).

van der Maas & Molenaar (1992) introduced catastrophe theory as a model to describe changes between qualitatively different stages of development. The main characteristic of the model is the existence of two qualitatively different behavioral modes, e.g. exemplar versus rule-based responding in categorization. As another example, in apparent motion perception, the two behavioral modes are perception of horizontal versus vertical motion of the dots (Ploeger et al., 2002). Catastrophe theory further describes the dynamics of changing between these behavioral modes as a function of so-called control variables, i.e. variables that control which behavioral mode is active. In the case of apparent motion, horizontal versus vertical motion perception is determined (among other things) by the aspect ratio of the rectangle at which the dots are placed.

In the current paper we focus on three specific predictions made by catastrophe theory. These three predictions follow from any model that incorporates multiple systems which generate qualitatively different behavior. The first prediction is bimodality, meaning that data are expected to be bimodally distributed. This really is the pinnacle of having two behavioral modes: behavior that is intermediate between the two modes can not and does not occur. For example, in categorization, people are assumed to use either an implicit or an explicit, verbalizable strategy (Ashby et al., 1998). At any one trial, only one of these strategies ultimately determines the behavior, and not a combination of the two. Second, catastrophe theory predicts divergence: bimodality increases as a function of one of the control variables. For example, in the case of attitudes, when people are highly politically involved, they tend to have stronger attitudes or opinions (positive or negative) in political matters than when they are less involved (van der Maas et al., 2003). Third, catastrophe theory predicts hysteresis, which stems from the persistence of behavioral modes. Hysteresis is the effect that switching from one behavioral mode to the other occurs at a different value of the control variable depending on the direction of the switch. For instance, in apparent motion, switches can be induced by changing the aspect ratio of the rectangle in which the points are organized. The aspect ratio at which participants switch from perception of horizontal to vertical motion is different from the aspect ratio inducing a switch in the other direction (Ploeger et al., 2002). For a more extensive description and discussion of catastrophe theory and its predictions, see van der Maas & Molenaar (1992).

Estimation methods and software are available for fitting the full catastrophe model to observed data (Grasman et al., 2009). However, fitting the full catastrophe model may be cumbersome and does not directly test the hypothesis of hysteresis as sequential dependency between data is not taken into account. As an alternative, here we propose to use constrained mixture and hidden Markov models to test the three predictions of catastrophe theory directly. The proposed method has wider applicability than to test catastrophe theory, as hysteresis (and divergence) are expected in other, more complex, dynamical systems as well. Whenever two stable modes of behavior exist, and there is, for example, negative feedback between the modes, hysteresis is exepected to occur.

Model

Central to catastrophe theory is the existence of a number of discrete, latent, cognitive states: e.g., we cannot observe whether someone is using an exemplar versus a rule-based strategy. Due to the latent nature of the states, we propose to use unsupervised learning techniques to infer their properties. In particular, we propose to use mixture and hidden Markov models, which allow us to infer the state-conditional distributions of the observed variables. By allowing the identity of a state to depend on (observed) covariates, we can further determine the dynamics of switching between the states.

In a mixture model, the probability density function f of a collection of observed variables $Y_{1:T} = (Y_1, \ldots, Y_T)$, can be written as:

$$f(Y_{1:T}|z_{1:T}, \boldsymbol{\theta}) = \prod_{t=1}^{T} \sum_{S_t} \mathsf{P}(S_t|\boldsymbol{\theta}_{\text{pr}}, z_t) f(Y_t|S_t, z_t, \boldsymbol{\theta}_{\text{obs}}) \tag{1}$$

where $S \in \{1, \ldots, N\}$ is a latent state or component, $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\text{pr}}, \boldsymbol{\theta}_{\text{obs}})$ is a vector of parameters for the state model and response model respectively, and $z_{1:T}$ are covariates. Note that the summation is over all possible realizations of $S_t \in \mathcal{S}$. For the response model, $f(Y_t|S_t, z_t, \boldsymbol{\theta}_{\text{obs}})$, we use generalized linear models, such that we can write the expected value (conditional upon the state), as e.g.:

$$E[Y|S = i, z] = g(\beta_0^{(i)} + \beta_1^{(i)} z) \tag{2}$$

where g is a link function relating the linear prediction to the expected value on the scale of the response variable Y.

A hidden Markov model can be viewed as an extension of a mixture model, in which transitions between mixture components (states) are modelled as a Markov process. In a hidden Markov model, the probability density function of the observed variables can be written as:

$$\begin{aligned} f(Y_{1:T}|z_{1:T}, \boldsymbol{\theta}) &= \sum_{S_{1:T}} f(Y_{1:T}, S_{1:T}|z_{1:T}, \boldsymbol{\theta}) \\ &= \sum_{S_{1:T}} \mathsf{P}(S_1|z_1, \boldsymbol{\theta}_{\text{pr}}) f(Y_1|S_1, z_1, \boldsymbol{\theta}_{\text{obs}}) \times \\ &\quad \prod_{t=2}^{T} \mathsf{P}(S_t|S_{t-1}, z_{t-1}, \boldsymbol{\theta}_{\text{tr}}) f(Y_t|S_t, z_t, \boldsymbol{\theta}_{\text{obs}}) \end{aligned} \tag{3}$$

where the sum is taken over all possible realisations of the state sequence $S_{1:T} \in \mathcal{S}^T$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\text{pr}}, \boldsymbol{\theta}_{\text{tr}}, \boldsymbol{\theta}_{\text{obs}})$ is a vector of parameters for the initial state model, transition model, and observation model respectively.

The main difference between a hidden Markov model and a mixture model (1) is thus the addition of a transition model, $\mathsf{P}(S_t|S_{t-1}, z_{t-1}, \boldsymbol{\theta}_{\text{tr}})$, used to predict the next state from the current state. In a mixture model, the states are assumed to be independent. In the remainder, we focus on models with two latent states, i.e. $\mathcal{S} = \{1, 2\}$. We can then use a binomial logistic regression model for the transitions:

$$\mathsf{P}(S_t = 2|S_{t-1} = j, z_{t-1}, \boldsymbol{\theta}_{\text{tr}}) = \frac{\exp(\eta_0^{(j)} + \eta_1^{(j)} z_{t-1})}{1 + \exp(\eta_0^{(j)} + \eta_1^{(j)} z_{t-1})} \tag{4}$$

Bimodality

Evidence for bimodality is obtained when a 2-state model fits better than a 1-state model (note that the power to detect this is larger when the state-dependent observation densities $f(Y|S, \boldsymbol{\theta}_{\text{obs}})$ are sufficiently different, e.g. $E[Y|S = 1] >> E[Y|S = 2]$).

Divergence

To test for divergence, we can make each state-dependent observation density dependent upon a control variable (the "splitting" variable). Generally, this means that the difference between the states increases as a function of the control variable. Using a generalized linear model (Equation 2), evidence for divergence is found when the slopes of the covariate are different, i.e. $\beta_1^{(1)} \neq \beta_1^{(2)}$, as this implies that the difference between the (state-conditional) expected values increases along the values of the covariate.

Hysteresis

Hysteresis means that behavioral states are persistent, and hence the history of the system partially determines the current state. Hence, a precondition for hysteresis is that the data are dependent (in time) rather than independent. In terms of a hidden Markov and mixture models, this means that a hidden Markov model should fit the data better than a mixture model, compared by appropriate goodness-of-fit statistics. Persistence of behavioral states means that the probability of remaining in that state is larger than the probability of leaving that state. In hidden Markov models, this means that $\mathsf{P}(S_t = j|S_{t-1} = j) > 0.5$. When a covariate or control variable has been measured, the prediction of hysteresis can be translated into a specific hypothesis concerning the transition model parameters. Hysteresis is found when the value of z such that $\mathsf{P}(S_t = 2|S_{t-1} = 1, z, \boldsymbol{\theta}_{\text{tr}}) = .5$ is different then the value such that $\mathsf{P}(S_t = 2|S_{t-1} = 2, z, \boldsymbol{\theta}_{\text{tr}}) = .5$. In the model of Equation 4, the point at which the transition probabilty is .5 is when:

$$P(S_t = j|S_{t-1} = i, z_{t-1}, \boldsymbol{\theta}_{\mathrm{tr}}) = .5 \longleftrightarrow z_{t-1} = -\frac{\eta_0^i}{\eta_1^i}$$

The condition for hysteresis is thus[1]:

$$\frac{\eta_0^{(1)}}{\eta_1^{(1)}} < \frac{\eta_0^{(2)}}{\eta_1^{(2)}} \tag{5}$$

Estimation

The parameters $\boldsymbol{\theta}$ in mixture and hidden Markov models are generally estimated with the Expectation-Maximization algorithm. This is an iterative algorithm which, starting from initial values of the model parameters, iterates through two steps until convergence is reached. In the Expectation step, the current values of the parameters are used to compute the expected log-likelihood function of the data (over all possible state sequences), given the current values of the parameters. In the Maximization step, this function is maximized to obtain new values for the parameters. Specification and estimation of general mixture and hidden Markov models is implemented in the R package `depmixS4` (Visser & Speekenbrink, 2010), which is used here to fit the models.

Example application I: testing bimodality in the balance scale task

The first example we present concerns a test of bimodality in a classical Piagetian task, the balance scale task. The substantive interest in these data is checking whether there are multiple modes of behavior versus the hypothesis that the performance differences are continuous variations of a single underlying mechanism (Inhelder & Piaget, 1958; Siegler, 1981; Siegler & Alibali, 2005).

The data consist of four items completed by 779 participants. The items are so-called distance items, in which the number of weights on each side of a balance beam is identical, and hence the distance to the fulcrum determines which side goes down. Figure 1 provides an example of such a distance item. Participants were presented with these items and asked whether the scale would tip to the left, tip to the right, or was balanced. The data used here are a subset of those presented and in Jansen & van der Maas (2002).

Here we analyze the total number of correct responses in four items; the frequencies of the sum scores are provided in Table 1, along with the mean age of the participants with that score. Note the high frequencies at 0 and 4 items correct, which suggests that participants either knew how to solve distance items, or they did not.

[1]When the inequality points in the other direction, this may be due to having two repellent states, rather than stable states.

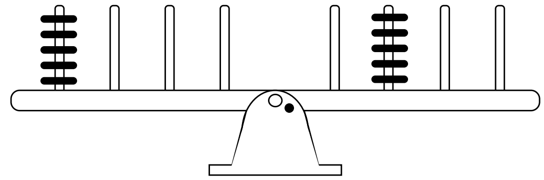

Figure 1: Balance scale distance item.

	0	1	2	3	4
participants	202	30	31	67	449
ages	9.1	10.9	10.6	11.9	13.1

Table 1: Sum scores of four items on the balance scale task and the mean ages of participants with that score.

The natural choice of distribution for these sum scores is the binomial, which is defined as:

$$f(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}, \tag{6}$$

where n is the total number of trials, p is the probability of success, and k is the number of successes, $k = 0, \ldots, n$. The number of trials n is a parameter of the binomial distribution that is taken to be known. In this example we have $n = 4$.

The model best suited to test the continuous variation hypothesis is a binomial (logistic) regression model, in which (the logit of) the probability of success p depends on one or more independent variables. Here we use age as independent variable, which is used in a binomial regression model.

Results show that age has a strong effect on the success probability $F(1, 777) = 345, p < 0.001$. E.g., according to this regression model, at age 6, the probability of success on these 4 items of the balance scale task equals 0.153, whereas at age 16 this probability equals 0.955.

The alternative model to test whether behavior is bimodal rather than unimodal, is a 2-state mixture model. As in the regression model, the component distributions are binomial; that is, the mixture components are characterized by different probabilities of success p.

The goodness-of-fit values of the regression model and the 2-state mixture model are provided in Table 2. For comparison, also a single state model is included. Table 2

model	LL	df	BIC
1-state	-1785.18	1	3577.01
2-state	-900.84	3	1821.66
regression	-1316.88	2	2647.07

Table 2: Goodness-of-fit statistics for 1 and 2 state binomial mixture models fitted on the balance data. For comparison, also the goodness-of-fit of a regression model is included with age as independent variable.

indicates that the 2-state model is the optimal model for these data. The parameters of the model are listed in Table 3 and they have a clear interpretation. The states represent correct versus incorrect strategies, respectively. Close to 70% of participants use a correct strategy for solving these items whereas the remainder uses an incorrect strategy.

state	P(state)	conditional
1	0.69	0.95
2	0.31	0.05

Table 3: Fitted parameters of the 2-state model of the balance data. 'Conditional' is the component specific probability of success for the binomial distribution.

Example application II: constrained mixture model for attitude data

This example concerns divergence that is predicted by catastrophe theory. This catastrophe flag means that the difference between behavioral modes increases as a function of the so-called 'splitting' variable of the model. To make this somewhat more concrete, we use an example from social psychology concerning the strength of attitudes held by people. The 'splitting' variable here is peoples' involvement with the issue at hand. When people are more involved with a particular issue, the strength of their attitudes is thought to increase, either in positive or in negative direction. When people are less involved, their attitudes are thought to be moderate. The data analysed concern a single attitude measured on a five-point scale from 'totally agree' to 'totally disagree'. Highly involved people have somewhat more extreme attitudes than those that are less involved. These data are from van der Maas et al. (2003) and they are available in R-package `cuspfit` by Grasman et al. (2009).

To test for divergence, we use a mixture model in which a participant's attitude is modelled as a binomial distribution. A comparison is made between two models; in both, models the responses are modelled as a binomial logistic regression model with strength of involvement (high or low) as a predictor.

The 1-state model, i.e. a simple binomial regression model, has an intercept of $\beta_0 = .35$ and slope of $\beta_1 = .09$. The latter is non-significant, $F(1, 1385) = 1.08, p > .2$.

The estimated 2-state model shows that in one state the effect of involvement on attitude is negative, while it is positive in the second state. In particular, at low involvement, both components have an average attitude of 2.93. At high levels of involvement, one component has an average attitude of 4.32, whereas the other component has 2.31.

The BIC statistics for these models are 4165 and 4162, respectively, indicating superior goodness-of-fit for the 2-state model. Hence, political involvement is a good

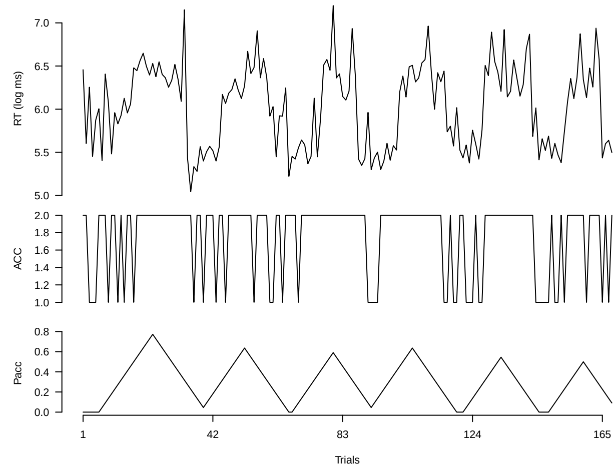

Figure 2: Response times (RT), accuracy (ACC) and relative pay-off values (Pacc).

'splitting' variable in the dynamics of attitude formation and increasing political involvement leads to polarization of attitudes.

Example application III: testing hysteresis in the speed-accuracy trade-off

This example concerns hysteresis in a lexical decision task under influence of a speed-accuracy trade-off. In the lexical decision-making task, participants were required to judge whether a combination of letters formed a word or not. Participants' behavior was manipulated by dynamically changing the reward that was paid for fast versus accurate responding. When $Pacc = 1$, the pay-off for accuracy was maximal, and there was no reward for speed; conversely, when $Pacc = 0$, the reward for accuracy was zero, and the reward for fast responding maximal. The data analysed here are from Participant A in Experiment 1 in Dutilh et al. (2011). Hysteresis was also tested in these data in Visser et al. (2009); the test for hysteresis used here is more general than was used there. During the task, $Pacc$ varied continuously in a zigzag manner and the participant was shown the reward after each trial such that behavior could be adjusted accordingly. The data consists of three separate series of trials, with 168, 134, and 137 trials respectively; Figure 2 shows the data of the first series.

A 2-state hidden Markov model was fitted to the data. In this model, responses Y were multivariate, including both accuracy (ACC) and response times (RT). These were assumed to be conditionally independent given the states. For ACC, a Binomial distribution was used, and for RT a Gaussian distribution; no covariates were included for these responses. However, $Pacc$ was assumed to affect the transition between the states, modelled as in Equation 4 with $z = Pacc$. The parameter estimates are given in Table 4. The parameter estimates clearly support the intended interpretation of the 2-state hidden Markov model, with state S_1 being the fast guessing (FG) state and state S_2 the slow and accurate stimulus-controlled (SC) mode of responding. The 2-state model (BIC = 564.87) clearly fitted the data better than a 1-state model in which RT and ACC were regressed on $Pacc$ (BIC = 870.01). Importantly, the parameter estimates clearly support the presence of hysteresis. According to the model, the probability of switching from the FG state to the SC state is 0.5 at $Pacc = 0.46$, while the probability of switching from the SC state to the FG is 0.5 at $Pacc = 0.21$.

state	$\mu(RT)$	$\sigma(RT)$	P(Ac)	η_0	η_1
1	5.52	0.20	0.53	-4.22	9.13
2	6.39	0.24	0.90	-3.37	15.80

Table 4: Parameter estimates of the 2-state HMM.

A simple condition for hysteresis is dependence in the data, which can be tested by comparing a 2-state mixture model with a hidden Markov model. Hence, next to the 1-, and 2-state HMMs, we also fitted a 2-state mixture model. Table 5 has the goodness-of-fit for the 2-state mixture model as the 'ind'epence model and the 2-state HMM as the 'hyst' model, as it is compatible with hysteresis. The BIC values clearly indicate that adding a transition model to the independent mixture model improves the goodness-of-fit substantially.

The condition of hysteresis (Equation 5) concerns inequality of the ratios of the intercepts and slopes. Hence, the third model we fitted is a model in which these ratios are constrained to be equal. This is model 'no hyst' in Table 5. To test whether the 'no hyst' model holds, a log-likelihood ratio test was performed which is reported in Table 5. The result is significant, $\chi^2(1) = 52.44$, $p < .001$, indicating that the ratio constraint on the parameters significantly decreases the goodness-of-fit of the model, and hence the data support the hypothesis of hysteresis.

model	LL	BIC	#par	χ^2	Δ#par	p
ind	-430.8	904.3	7			
hyst	-249.2	553.2	9			
no hyst	-275.4	599.5	8	52.4	1	0

Table 5: Goodness of fit statistics for speed-accuracy data models, see text for details.

Conclusion & Discussion

Catastrophe theory makes very specific predictions about the dynamics of switching between different cognitive states, and the resulting behavior. Here we have shown how to test three essential predictions made by catastrophe theory: bimodality, divergence, and hysteresis, in three different data sets. Two data sets contained cross-sectional data which does not allow to model state transitions. Using mixture models, we could still test the catastrophe flags of bimodality and divergence for these data. Clearly, these are necessary, but not sufficient conditions for a dynamic systems model such as the catastrophe model. Testing predictions about state transitions derived from these models requires sequential observations from a single unit, such as in the last example, where we tested for hysteresis with a hidden Markov model.

The framework of catastrophe theory provides a strong foundation to study the dynamics of changing behavior by neccessitating the identification of control variables. In social cognition, one of the control variables was political involvement. A second important control variable is political orientation: right- versus left-wing van der Maas et al. (2003). In the balance scale task, younger children focus on just the number of weights on either side of the balance beam and ignore the distance, while older children tend to rely on both dimensions to decide which side goes down, although typically they have trouble integrating these dimensions correctly (Jansen & Van der Maas, 2001; Jansen & van der Maas, 2002).

Hysteresis is a particularly interesting phenomemenon predicted by dynamical systems models, such as catastrophe theory. Being able to straightforwardly test this prediction, along with bimodality and divergence, provides the possibility of applying this framework in other areas as well. We already mentioned categorization learning: there clearly are two behavioral modes, exemplar versus rule-based, that are associated with particular behavior. The variables that control whether one or the other type of behavior occurs are to some extent known from earlier work. Bimodality of behavior is found for example in Ell & Ashby (2006), as a result of varying the degree of overlap between categories to be learned. Developmental psychology offers a wealth of examples where catastrophe theory can be applied. An interesting example concerns the dimensional change card sorting task, where lack of inhibition of prepotent responses prevents children from switching to following novel instructions (van Bers et al., 2011). Memory development using mixture models is studied in Koppenol-Gonzalez et al. (2013) and reasoning in Bouwmeester et al. (2007).

In sum, we conclude that catastrophe theory provides a strong theoretical framework for studying the dynamics of switching between competing cognitive states and their resulting behavior. Having straightforward tests for a number of predictions made by catastrophe theory by using mixture and hidden Markov models as we have presented here, makes this framework available to a larger group of researchers.

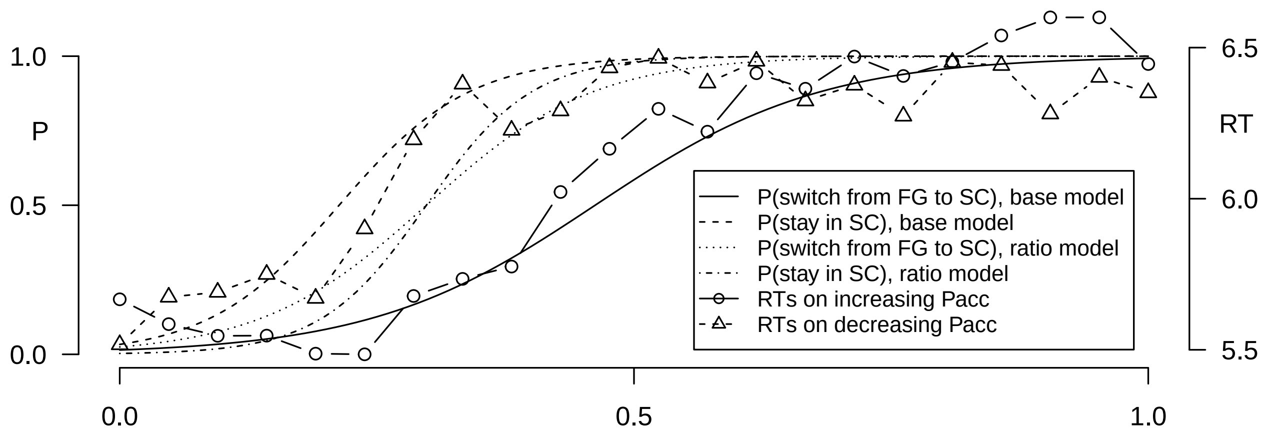

Figure 3: Transition probabilities as a function of *Pacc* for the hysteresis versus the no-hysteresis model; included are the mean RTs on trials with increasing versus decreasing values of *Pacc*.

References

Ashby, F., Alfonso-Reese, L., Turken, A., & Waldron, E. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–482.

Bouwmeester, S., Vermunt, J. K., & Sijtsma, K. (2007). Development and individual differences in transitive reasoning: A fuzzy trace theory approach. *Developmental Review*, *27*, 41–74.

Dutilh, G., Wagenmakers, E.-J., Visser, I., & van der Maas, H. L. J. (2011). A phase transition model for the speed–accuracy trade–off in response time experiments. *Cognitive Science*, *35*, 211–250.

Ell, S. W., & Ashby, F. G. (2006). The effects of category overlap on information-integration and rule-based category learning. *Perception & psychophysics*, *68*(6), 1013–1026.

Grasman, R. P., van der Maas, H. L., & Wagenmakers, E.-J. (2009). Fitting the cusp catastrophe in r: A cusp-package primer. *Journal of Statistical Software*, *32*(8), 1–28.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Jansen, B. R., & Van der Maas, H. L. (2001). Evidence for the phase transition from rule i to rule ii on the balance scale task. *Developmental Review*, *21*(4), 450–494.

Jansen, B. R. J., & van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, *81*(4), 383–416.

Johansen, M., & Palmeri, T. (2002). Are there representational shifts during category learning. *Cognitive Psychology*, *45*, 482–553.

Koppenol-Gonzalez, G. V., Bouwmeester, S., & Vermunt, J. K. (2013). Short term memory for serial order: unraveling individual differences in the use of processes and changes across tasks. *Frontiers in psychology*, *4*.

Ploeger, A., Van Der Maas, H. L., & Hartelman, P. A. (2002). Stochastic catastrophe analysis of switches in the perception of apparent motion. *Psychonomic Bulletin & Review*, *9*(1), 26–42.

Schmittmann, V. D., Dolan, C. V., van der Maas, H. L. J., & Neale, M. C. (2005). Discrete latent Markov models for normally distributed response data. *Multivariate Behavioral Research*, *40*(4), 461–488.

Siegler, R. S. (1981). *Developmental sequences within and between concepts*. No. 46 in Monographs of the Society for Research in Child Development. SRCD.

Siegler, R. S., & Alibali, M. W. (2005). *Children's Thinking*. Upper Saddle River (NJ): Prentice-Hall.

van Bers, B. M. C. W., Visser, I., van Schijndel, T. J. P., Mandell, D. J., & Raijmakers, M. E. J. (2011). The dynamics of development on the dimensional change card sorting task. *Developmental Science*, *14*(5), 960–971.

van der Maas, H. L., & Molenaar, P. C. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological review*, *99*(3), 395.

van der Maas, H. L. J., Kolstein, R., & van der Pligt, J. (2003). Sudden transitions in attitudes. *Sociological Methods & Research*, *32*(2), 125–152.

Visser, I., Raijmakers, M. E. J., & Van der Maas, H. L. J. (2009). Hidden markov models for individual time series. In J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra, & N. Chaudhary (Eds.) *Dynamic Process Methodology in the Social and Developmental Sciences*, chap. 13, (pp. 269–289). New York: Springer.

Visser, I., & Speekenbrink, M. (2010). depmixS4: An R-package for hidden Markov models. *Journal of Statistical Software*, *36*(7), 1–21.