

# Dynamic Modeling for Development and Education: From Concepts to Numbers

Paul Van Geert<sup>1</sup>

**ABSTRACT**— The general aim of the article is to teach the reader how to transform conceptual models of change, development, and learning into mathematical expressions and how to use these equations to build dynamic models by means of the widely used spreadsheet program Excel. The explanation is supported by a number of Excel files, which the reader can download at [www.paulvangeert.nl/articles\\_appendices.htm](http://www.paulvangeert.nl/articles_appendices.htm). The article begins with a discussion of how to define variables in the context of modeling and the difference with the context of measurement. It proceeds with a simple dynamic model to give the reader a general idea, and continues with an explanation of the extended logistic model, which can be conceived of as a building block of more complex growth models pertaining to learning and development. The final section discusses the building of a concrete developmental model, answering the question of why girls like pink.

The issue of dynamic modeling of developmental and educational processes can be approached from two different sides. Abraham (forthcoming) approaches modeling from the general theory of dynamical systems, working his way down to the actual empirical phenomena and the theories that explain them. In this article the opposite approach will be followed. I will start from processes of learning and development and the conceptual theories or principles aimed at explaining them, and will then go on with the question of how these conceptual principles can be transformed into mathematical equations. Finally, a concrete tutorial will be given on how dynamic models can be constructed by means of standard spreadsheet software (Microsoft Excel).

<sup>1</sup>University of Groningen

Address correspondence to Paul Van Geert, The Heijmans Institute, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands; e-mail: [paul@vangeert.nl](mailto:paul@vangeert.nl)

## VARIABLES, MEASUREMENT, AND MODELING

In 1946, Stanley Smith Stevens published a four-page article in *Science* that would have an enormous influence on empirical research in psychology (Stevens, 1946). Stevens sought to answer the question whether it is possible to measure human sensation. He defined measurement as the assignment of numbers to objects or events according to rules. Different kinds of rules would lead to different kinds of scales and to different kinds of measurement. In his article, he suggested the now well-known distinction between levels or scales of measurement: the nominal scale, the ordinal scale, the interval scale, and the ratio scale. Most of the modeling that we do in science also concerns the assignment of numbers to objects or events according to rules. In that sense, measurement and modeling have something very important in common. However, modeling is not the same thing as measurement. Suppose that we would like to model the behavior of a student in primary school during a math lesson (e.g., see Steenbeek & van Geert, 2013; this model was based on a dynamic model of dyadic play that focused on alternations between playing together and playing alone; Steenbeek & Van Geert, 2007, 2008). For each second during a math instruction, the model generates a categorical variable, which is either on-task or off-task behavior. However, the average duration of on-task behavior per unit time, such as a minute or an entire math lesson, amounts to a ratio variable, which can be directly observed and compared with the output of the mathematical model.

In their agent-based dynamic model, Steenbeek and Van Geert (2013) sought to understand changes in on- and off-task behavior by referring to the concerns or interests of the agents, namely the child and the teacher. Every agent has concerns that emerge in a particular context and that are based on general bio-social concern structures, described in Deci and Ryan's (2009) theory of self-determination. This theory makes a distinction between the basic concerns of autonomy, competence, and relatedness. In addition to the basic concerns,

the agent-based dynamic model assumes that there are two important context-specific concerns or “interests” in the math lesson, namely the concern for doing the math assignment and the concern for doing something else. The model treated these two concerns as ratio variables, which featured in the model equations. However, the difficulty is that the authors could not measure these concerns, let alone that they could measure them in the form of ratio variables. Maybe in the near future, researchers will be able to perform this sort of measures by means of nonintrusive techniques, such as optical brain imaging *in vivo* (Hillman, 2007). At this particular moment, however, we don’t have the methodological tools to actually measure those two concerns as they emerge and fluctuate during this particular math assignment. Does that mean that we need to postpone the modeling of the dynamics of these concerns until an adequate form of measurement is available?

Fortunately as I stated earlier, measurement and modeling are different things. As ratio variables are the most important type of variable in dynamic modeling, it is important that we can *theoretically* assign ratio properties to the variables we work with in our models. This theoretical or conceptual question is different from the question of how to assign ratio properties to the variables we *empirically* observe or measure.

How shall we conceptually or theoretically assign numerical ratio properties to variables such as “context-specific on-task concern” and “context-specific off-task concern”? We can do so if the following three properties can be meaningfully assigned to these variables. First of all, the variable must, in principle, have a true zero value. That is, it must be conceptually feasible to think of a situation in which the on-task concern, for instance, is zero. Second, it must be conceptually feasible that these two variables can be treated as one-dimensional variables. That is, although in reality a specific concern might be determined by a wide variety of different influences, conceptually speaking, specific concerns can be treated as something that varies along one dimension (Van Geert, 1998). Third, it must be conceptually feasible that the variable can undergo continuous change (which implies an underlying continuous dimension). For instance, it must be (conceptually) possible for the concern to increase or decrease in strength or intensity in a continuous way and to do so with continuously varying change rates. These three properties can be conceptually assigned to a variable such as concern strength, or to host of other variables, such as the amount of support a particular family gives to the interest in science of one of its children (Van Geert, Den Hartigh, Steenbeek, & Van Dijk, submitted). In principle, this support may be zero, one-dimensional, and continuously changing in intensity (including remaining stable over time). Compare this family support variable with another family-related educational variable, namely parenting style. Parenting style is a categorical variable, hence, it cannot be zero and it is not one-dimensional. However, the frequency with which

a particular parenting style, such as authoritative parenting, occurs in a particular family over a particular period of time, can in principle be treated as a ratio variable, and thus be subject to typical dynamic modeling. The assumption that the conceptual variables in dynamic models must be typically defined on a ratio scale in general implies that these variables will be more or less domain-specific. Intelligence, for instance, is a very general variable, and it is hard if not impossible to think of an organism with zero intelligence. However, it does make sense to assign a zero size lexicon to a baby, or a zero size numerical addition skill (although there is research showing that babies have very elementary number abilities, see, e.g., Dehaene, 2007).

## TYPES OF DYNAMIC MODELS

Dynamic models can be written in various formats. To begin with, you can specify a dynamic model in a purely conceptual form, by verbally describing the nature of the variables, of the principles that govern the *change* of these variables, and of the dynamic connections between variables that contribute to the change of either of them. Second, you can specify a dynamic model in a mathematical form. The mathematical equations specify the *change* in the variables by means of mathematical expressions that correspond with the theoretically specified forms of change and the associations between the variables. The mathematical equations represent causal relationships, that is, how one variable causes another variable to change. Thirdly, you can specify a dynamic model in the form of a particular computer model that allows you to study the model by running simulations based on the mathematical equations and the various values of your parameters (see Abraham, forthcoming; see further in this article).

There exist various ways of transforming conceptual and/or mathematical models into computer models. The Net Logo software used in Abraham’s article is an example of a particular approach to modeling, which is based on the notion of agents (see also Steenbeek and Van Geert’s [2008, 2013] agent-based models of dyadic play and math lessons). Many dynamic models, however, typically focus on a collection of variables that interact within a single individual. For instance, if a researcher wishes to understand the relationship between the developmental trajectory of representational mappings and that of representational systems in a particular child (e.g., Fischer & Bidell, 2006), the topic of analysis is a system of two interacting variables in a single subject. Another example is from a recent study by Van Dijk et al. (2013) on the dynamic relationship between the number of words (MLU) in a mother’s child-directed speech and the number of words (MLU) in the child’s speech. This is a system of two interacting variables in a dyad. There exist various types of software that allow the modeler to build a dynamic model of the interaction and change of these two variables. For

instance, *Stella* (discussed in Abraham's article) is a typical example of such a software. *Stella* and comparable software such as *Powersim* or the freeware program *Insight Maker*, are based on two fundamental types of components, namely *stocks* and *flows*. This terminology typically stems from business science, in which stocks of goods are filled or emptied. Thus, there is a "stock" of words to be used in a single sentence by the child and one by the mother, and there are flows of the number of words (MLU), such as an increase in those of the child and a temporal decrease in those of the mother. Behavioral scientists would use the word variable instead of stock, and change instead of flow.

#### SOME CONVENTIONS REGARDING SPREADSHEETS AND MODELING

A spreadsheet is typically divided into rows and columns. Columns are represented by capital letters A, B, up to XFD, and rows are represented by numbers (1-1,048,576). Names of rows and columns cannot be changed, but you can easily add a new name to ranges consisting of columns and rows (see further).

In my examples, I shall refer to column A, column B and so forth, or to cells A1 and B1 for instance, but the reader is free to use other columns, use a new worksheet and use different cells. What is important, however, is that the relationship between cells is respected. Suppose that I say that a model starts in cell A1 and that cell A2 must refer to cells A1 and B1. The reader might have decided to start his or her model in cell E11 for instance. If that is the case, the readers' cell E12 must refer to cells E11 and F11.

Spreadsheets use standard symbols for mathematical operations:

\* means "multiply"

/ means "divide"

^ means raise "to the power of"

A typical spreadsheet expression, for instance one that appears in cell A2, might look a bit like this:

$= (A1 * 2) + B1$

which literally means: the content of cell A2 is equal to the content of cell A1 times two, plus the content of B1.

Spreadsheet expressions can also contain logical expressions, and a typical one for instance appearing in cell B1 might look like this:

$= \text{If}(A1 > 10, 1, 0)$

which literally means: if the content of cell A1 is bigger than 10, then the content of cell B1 is 1, if A1 is not bigger than 10, B1 is 0.

Logical and numerical expressions can be combined, for instance in cell B1 you might find the following expression:

$= \text{If}(A1 > 10, 4, 2) * A2$

It means that if the content of A1 is greater than 10, the content of B1 equals 4 times the content of A2, else, the content of B1 equals 2 times the content of A2.

In summary, a spreadsheet program such as Excel contains a relatively simple but powerful programming language, with which growth models can be constructed.

Since I started with dynamic modeling in the field of education, learning, and development, which goes back to 1990 (see Van Geert, 1991), I have deliberately chosen a different representational format than the stock-and-flow models. The modeling format that I have chosen is that of a spreadsheet program, the most widely used example of which is Microsoft Excel. Strictly speaking, it has mathematical limitations (all models are maps; see Abraham, forthcoming), but in practice the limitations will not hamper the researcher from using Excel as a convenient modeling tool. In addition, it has some interesting didactic advantages.

If you open Excel, you see that it is divided into rows and columns, which are numbered by integers and letters respectively (see Textbox at left). Let us take one such column and conceive of it as a sequence of moments in time. That is to say, every cell in the column is a particular moment in time, and the next cell (indicated by the next row in the spreadsheet) is simply the next step in time. You can define "step" pragmatically in accordance with what you wish to model: a second, an hour, a day, a week, and so forth. For instance, if you define a cell in a particular row as a day, a column of 365 such cells will constitute a year. Instead of making a cell correspond with a particular moment in time, you can also make it correspond with a particular event, such as a particular math lesson. If there are eight math lessons every week in a school year that counts 40 school weeks, a sequence of 320 cells corresponds with an entire school year.

#### COPYING AND PASTING IN SPREADSHEETS

One of the convenient features of spreadsheets such as Excel is the flexibility of copying and pasting. It is important to know that you can copy the content of the cell either as a *formula* or as a *fixed content*.

Suppose cell A1 contains the value 2, and cell A2 contains the formula  $= A1 * 2$ , which cell A2 will show as a value, namely 4. First copy cell A1. You can do so by clicking on cell A1, then use the control-C key combination (in Windows). You can also click on cell A1 and, under the home menu tab click on the Copy menu item. In order to paste the formula content of A1 to the range

of cells spanning A2–A60, first select the range A2–A60, then use a key combination Ctrl-V (Windows). By doing so, the mathematical relationships between the cells are kept intact. For instance, if you click on cell A30, it will contain the formula = A29\*2; and cell A45 will contain the formula = A44\*2. Note that the cells themselves will show you numerical values, but if you click on a cell, you will see the formula content of the cell in the *formula bar* which appears on the right on top, under all the menu items.

If you have made an interesting model simulation, and you want to keep the numerical values of the model in a separate place or worksheet, you should paste the numerical and not the formula content of the cells. Select the whole range of cells corresponding with your model, use Ctrl-C or the Copy menu item, and then click on the cell where you want your range of pasted numerical values to start. Then click on the Paste item (on the left, under the Home menu tab), then click on the Paste Values menu item.

An interesting feature of the spreadsheet is that a cell can have a value that depends on another one. For instance, if cell A1 has a value 0.5, and in cell A2 you type = A1\*2, the value of cell A2 is 1, which is  $0.5 \times 2$ . By copying a cell to the next one in the column, you actually copy the way a cell refers to another cell, and not the value of the cell. If you copy the content of cell A2 and paste it into A3, you see that the content of A3 = A2  $\times$  2. If you copy the content of the second cell to all the remaining cells in a column that spans cell A1 to cell A64, and you treat every cell as a step in a sequence of events, you have actually modeled a simple dynamic process, which is the doubling scenario (every value is twice the value of its preceding cell, meaning that every next step in the process is twice the value of the preceding step). You can select these 64 cells, then click on the menu item Insert, then click Line, then click on the line graph type of your choice, then click on a place in the worksheet where you want the graph to appear. You just built your first dynamic model.

The structure of this article is as follows. To introduce the reader to the world of developmental modeling in spreadsheets I shall begin with a very simple model of change. The model uses the concept of random influences, and aims to show how a simple conceptual model of change in children's motivation can be transformed into a sequence of mathematical equations characteristic of dynamic modeling in general. I will then proceed by introducing *growth modeling*, and focus on the model of logistic growth, which is more or less the “workhorse” of dynamic growth modeling. The logistic model can be used in a wide variety of developmental models and I will give a short overview of the models that have been built on this framework. Finally, I will apply what we have learned about modeling to

the question of why girls like pink. Answering this question via dynamic modeling, I will illustrate a variety of modeling principles that readers can use to start working on their own developmental or educational modeling questions.

### A SIMPLE MODEL OF CHANGE AS A STARTING POINT

Let us do the following *Gedanken experiment*. Suppose we have an absolutely valid test to measure the intrinsic motivation of a particular child for making the math assignments in class. Suppose that today, we have measured the child's intrinsic math motivation, and found a score of 100. Intrinsic motivation is not fixed; it may increase or decrease, tomorrow and the day after tomorrow, and so on, dependent on the child's experiences with a particular content (see, e.g., the studies by Hidi & Renninger, 2006, on the development of interest). For simplicity, we assume that *changes* in intrinsic motivation depend only on the child's concrete, event-based experience. Let us treat a math assignment as an event that increases the child's intrinsic motivation if the child's evaluation of that assignment is globally positive and decreases the intrinsic motivation if the evaluation is negative. Let us assume that *on average* the child's evaluation is neutral, which actually means that in the long run there are about as many positive as negative evaluations, which will cancel each other out. In our model, we shall assume that the child's evaluation of a particular math assignment increases or decreases the current intrinsic motivation with a randomized magnitude of change. That is, the change has an average of 0 and a standard deviation of 1, randomly drawn from a normal distribution. It is important to note that this model refers to individual children: it will generate developmental trajectories each of which refers to an individual case. The principle of change featured by the model is assumed to be general, that is, applying to every possible individual case. This is quite different from standard models in the literature, which mostly describe developmental trajectories that refer to the population, that is, average trajectories.

We can model these individual change processes as event series in Excel (see Figure 1). Suppose cell A1 (or any other cell you choose) represents the score of a particular child the first measurement, in which case we type 100 into cell 1. Excel provides a formula that allows us to calculate a random number drawn from a normal distribution, with given mean and standard deviation (which are 0 and 1 in this particular case). Here's the equation:

$$= \text{NORM.Inv}(\text{rand}(), 0, 1).$$

To draw this random number, Excel uses another random function *rand ()*, which draws a random value between 0 and 1. The random number is the first argument in the equation, the second argument is the average value which is zero, and

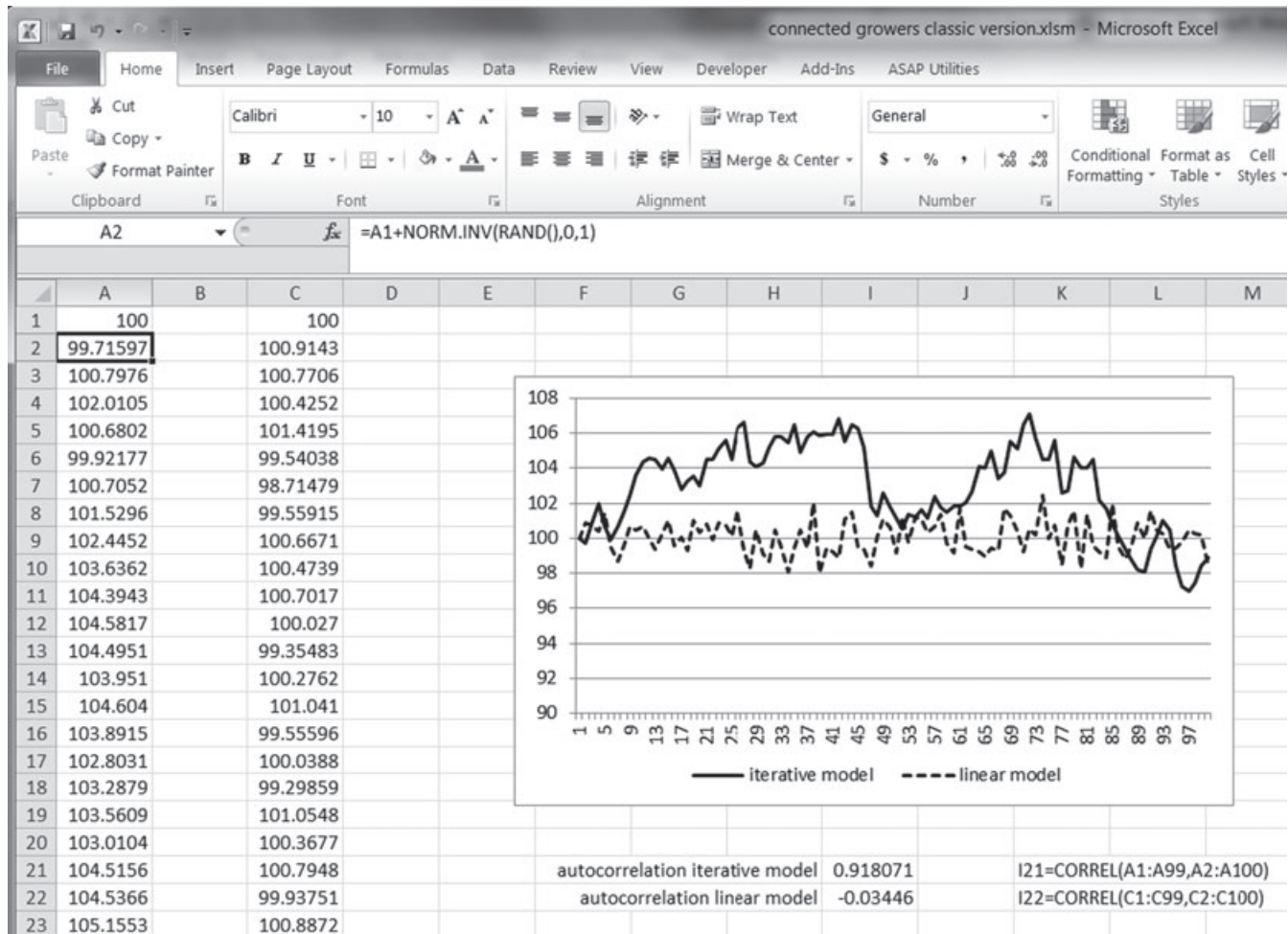


Fig. 1. Screen picture of the Excel file modeling the change in motivation, based on the assumption that the average change is represented by a random number with mean of zero and standard deviation 1. The dynamic (iterative) model shows typical long term “swings,” which explain the high autocorrelation.

the third argument is the standard deviation which is 1. It goes without saying that the value of these numbers can be freely chosen by the model builder. According to this model, the intrinsic motivation for math during the second assignment is equal to the intrinsic motivation during the first assignment, plus the positive or negative effect of the first assignment. To write this relationship in a mathematical form, click on cell A2, which represents the second math assignment, and type

$$= A1 + \text{NORM.INV}(\text{rand}(), 0, 1)$$

Next, repeat this equation in all the cells up to cell A100. To do so, copy the formula to cells A3–A100, which simulates a sequence of 100 math assignments made by this particular child. In every cell, we will find the result of the preceding cell with a random number added or subtracted from it, and drawn from the same normal distribution. Each time you click on the F9 function button, Excel will recalculate the model with new random values for each of the cells (Mac users should try

CMD+= (equal sign) or F9 to calculate all sheets in all open workbooks, or SHFT+F9 to calculate the active worksheet). Insert a line graph representing cells A1–A100 to facilitate the inspection of the modeling results (to do so, select the range from A1 to A100, select Insert from the menu bar, then select Line Graph, then select the type of line graph you wish to use, then select a cell where your line graph should appear).

The reader might think that, since we add 100 random numbers with an average of zero, the net effect of the simulated 100 process steps will also be close to zero, since it is very likely that these random numbers will cancel each other out. Under this assumption, we expect that after 100 steps the child will still be quite close to its starting level of motivation, which our test measured as 100. However, the temporal trajectories generated by our simple model are very different from this scenario. In a number of cases we will see that motivation goes up or goes down most of the time; in other cases, we will see that it goes up quite significantly and then moves back

towards its starting point, and eventually shows a significant downward trend. The majority of these trajectories, which the reader can very easily inspect (by pressing the F9 button, which will recalculate the model by recalculating the random numbers) will considerably deviate from the “flat” case that tends to vary a little around the value of 100. This result illustrates a very important feature of dynamic models, namely that they are *iterative* (one can also call them recursive). An iterative or recursive model takes the preceding value as its input, applies some form of change to this preceding value—in this particular case adding a random number—and produces an output which will serve as an input for the next step in the process. The iterative or recursive nature of the application of the mechanism of change is a fundamental feature of dynamic models (Van Geert & Steenbeek, 2005a, 2005b; Weisstein, 1999), which makes them very different from the linear regression model that most researchers in the educational and developmental sciences will be used to.

#### LOGISTIC GROWTH AS A GENERAL MODEL OF DEVELOPMENTAL CHANGE

In the 19th century, demographers and statisticians were greatly concerned with the growth of populations. Thomas Robert Malthus (1766–1834), for instance feared that the human population would grow far beyond the available food and energy resources, and maybe become extinct. The Belgian astronomer-demographer-statistician François Ferdinand Verhulst (1804–1849), however, discovered a more realistic model of population growth, which among others allowed him to relatively accurately predict the population of Belgium in the 20th century. This model was called the *logistic model*. This word refers to the French word “logis,” which means the housing and associated catering of military troops. However improbable it may seem, especially in light of its slightly military origin, the logistic model turned out to provide a very general conceptual model for growth models of virtually any kind, including developmental growth, cognitive growth, skill development, education, and so forth. In what follows, I shall try to develop the conceptual basis of the logistic growth model and show how it may be written in the form of dynamic models in Excel.

In order to arrive at the logistic model, we will first go back to the simple model of motivation change described in the preceding section. In fact, if you write it up in a different notation, the motivation model we have built in that section amounts to the following form of change:

$$\Delta M / \Delta t = a.$$

This equation means that the change  $\Delta$  in some variable  $M$ , for instance motivation (or reading comprehension, addition and subtraction skills, etc.) over some particular period of

time written as  $\Delta t$ , is equal to a value  $a$ , which in this case is a random number drawn from a normal distribution with 0 mean and standard deviation of 1. That is to say, for every “tick” of the developmental clock, we add or subtract a certain amount of motivation (for instance) to the motivation that is already there.

However, from the viewpoint of growth and development, there is a serious theoretical problem with this additive model: it says that the amount of change is completely independent of what is already there. This principle runs against everything we know about development. Processes of development are almost by definition determined by what is already there, by the amount of understanding or skill already achieved by a child as a result of the preceding developmental processes. It was one of the fundamental discoveries of Piaget that children developed because they acted, interpreted, and evaluated in terms of the developmental structures they already possessed. It was a fundamental discovery of Vygotsky that educators present children with help and instruction based on what the children already understand and can successively do with the help of the educators.

#### Step One: Exponential Change

The simplest mathematical way for expressing the fact that the amount of change is a function of what is already there is by means of a simple multiplication: we express the change in a particular variable, such as a child’s current reading comprehension, as a form of scaling (for a discussion on the meaning of multiplication as a form of scaling, see for instance Devlin, 2013).

Suppose that the amount of change per month is specified by a change parameter called  $r$ . In that case, the scaling factor per month is equal to  $1 + r$ , that is, to say after a month the current level will have been scaled up to  $1 + r$  of its current value.

We call this model the *exponential growth model*, and it is the first step towards the logistic model. In terms of a simple Excel model, you might do the following: in cell A1 write the value of a starting point, for instance

$$= 0.01$$

and in cell A2 write

$$= A1 + A1 * 0.05$$

and in cells A3–A50 copy the content of cell A2 The growth curve produced by this model shows a pattern of acceleration. If our time series is long enough, such models arrive at impossibly high values (it was the idea of unlimited growth that scared Malthus to death, metaphorically speaking). That is, there is nothing in the model that can stop it. From our observation of developmental and educational processes we know that real processes tend to arrive at more or less stable

levels, which are indeed greatly dependent on individual and environmental differences. In order to explain the stability, we can invoke some sort of external factor that stops the growth. For instance, learning can be stopped by the cessation of teaching. However, under the assumptions of the exponential model, if one goes on teaching forever, learning will go on forever, with ever-increasing returns. This does not seem to be a realistic representation of how learning, development, or teaching, actually work. The model of restricted growth, which we will now explain, tries to solve the problem of unlimited growth in the following way.

### Step Two: Restricted Change

Developmental processes often involve abilities, skills, or knowledge that are formed based on a particular goal, example, or model in the literal sense of the word. For instance, language acquisition implies that children acquire a particular language spoken in their environment, with a finite lexicon and grammatical rules. When children learn to ride a bicycle, the nature of the skill is well defined in advance: you have to keep yourself in balance on the cycle, avoid obstacles, and move at a certain speed (of course, as soon as you master the basic skill, you can further improve it up to the point that you can win the Tour de France, for instance). That is to say, developmental progress can be a function of what you still cannot do in the perspective of a particular, predefined goal state. Assume that we have a developmental ruler expressing a particular ability, and the goal state you can perceive as a learner is characterized by a level  $L_G$ .<sup>1</sup> Your current level is characterized by the level  $L_t$ . The amount of progress you make over a particular period of time is a function of the distance between the two, namely of  $L_G - L_t$ . More generally, your progress amounts to scaling up your current level by a factor equal to

$$1 + r(L_G - L_t).$$

You can easily see that as  $L_t$  approaches  $L_G$ ,  $L_G - L_t$  approaches zero, thus reducing the amount of change to zero, thus leading to stability. We call this particular model the *restricted growth model*, because it is restricted by a pre-set goal level. It is yet another step in the direction of the logistic growth model, and we shall see why neither the exponential nor the restricted growth model alone can cover developmental growth processes.

Nevertheless, the restricted growth model can provide a dynamic explanation for relatively simple processes of learning, which were extensively studied in the classical learning theory literature (e.g., Walker, 1996).

For instance, suppose we wish to teach the names of the countries forming the Eurozone, or the names of the states forming the United States, to a child who does not know any of these names. These names clearly form a finite list (18 countries and 50 states respectively). To model this type of

learning process, take cell A1 has your starting point, and in A1 write

$$= 0,$$

which is the number of names of countries or states known by this particular child before the teaching process has taken place. Let us say we practice the names of the states every day, and we shall assume that the learning rate of a particular child is equal to 0.1 per day. In cell A2 write

$$= A1 + 0.1 * (50 - A1),$$

and copy this to cells A3–A50. The reader can check the model and see that, during the first day, the child has learned the names of five states. However, at day 19, the child has learned 0.67 names of states, which is indeed a big difference with the first day. In practice, learning 0.67 names a day is a way of saying that the child has about 67% chance that it will have learned yet another name of a state, and 33% chance that it has learned anything at all. In fact, the model very clearly shows that under a similar effort, namely a similar practice every day, the returns are greatly diminishing as the learning process lasts longer.

Before making the final step towards the logistic growth model, we should note that whereas the restricted growth model can start from a zero level, the exponential growth model must always have something to begin with. That is to say, there must be some seed or starting point for it. In developmental processes this seed or starting point is likely to emerge on the basis of processes or skills that are already present. For instance, in the unlikely event of a child who has never seen any written words in his life, the first confrontation with a written word at school will trigger some sort of seed of reading comprehension in the form of general processes of visual perception and form recognition in the child, which the teacher can explicitly associate with a particular sound pattern. This seed of reading comprehension, which is clearly the product of the interaction between the child and the instructor, can form the starting point of a successive process of scaling up the reading comprehension skill.

### Step Three: Combining Exponential and Restricted Change Into the Logistic Model

A child constitutes a developmental system, which consists of many skills, abilities, psychological functions, material tools, social supports, instructors, etc. What would happen if some of the abilities, skills, or psychological functions followed the *exponential* growth trajectory? Exponential growth diverges into ever-increasing numbers. However, the developmental system is a system characterized by physical limitations. Take for instance the brain. All skills and abilities correspond with particular neural networks of connections in the brain. Increasing specialization in general requires increasing neural

networks. In the limit, it is thinkable that the exponentially growing abilities will start to compete for “brain space,” and that this competition will have a negative effect on their rate of growth. On the other hand, we know that as abilities or skills are expanding into exceptional forms, such skills can only be maintained if they are constantly practiced (think for instance about exceptional musical skills or sports skills). That is, there will be a moment at which such exponentially increasing skills will start to compete for the limited amount of time a person has available to practice and maintain them (a day has only 24 hours). Given enough skills and abilities that require practice, the system will very soon reach the limits of its available resources, and this will begin to have a negative effect on the—so far exponential—growth rates. That is to say, in a developing system, or in any system of growers for that matter, the rate of growth of a particular grower will decrease as the level of that grower increases, due to the competition for limited resources (brain resources, time, external support, etc.) that are characteristic of any physical system. That is to say, the growth parameter  $r$  from the exponential model of a particular skill is not a constant, but a variable, which we shall now represent by the capital letter  $R$ .

How can we express the fact that  $R$  becomes smaller as the variable, whose growth it determines, becomes bigger? A simple way to express this fact is by rewriting  $R$  as the difference between a constant growth rate (a parameter we shall again call  $r$ ) minus some scaled version of the current level of the variable,  $a * L_t$  ( $a$  is the scaling parameter for  $L_t$ ). The value of the scaling parameter typically results from the interactions of all the components in the system.

We can thus write a new version of the original exponential growth model as follows:

if  $\Delta L / \Delta t = L * R$ , then  $R = r - a * L$ , from which follows that

$$\Delta L / \Delta t = L * R = L (r - a * L)$$

In this equation growth will come to a halt at the moment where  $a * L$  has become similar to  $r$ , because then  $r - a * L = 0$  and  $\Delta L / \Delta t$ , which is equal to  $L * 0$  is 0. What is the value of  $L$  for which  $\Delta L / \Delta t$  is 0? It is a value for which  $L (r - a * L) = 0$ . Since  $L$  cannot be 0 (see the discussion above), it is also the value for which  $r - a * L = 0$ , and thus it is a value for which  $L$  equals  $r/a$ . This is the value which ecological models have termed the *carrying capacity of the system*. That is to say, the entire system, which is a developmental system in this particular case, can “carry” or maintain a particular level of  $L$ . This level is equal to  $r/a$ , a ratio which is typically represented by the constant  $K$ , which stands for carrying capacity. Hence, this constant  $K$  is a property typically associated with the particular variable  $L$  and with the way all other variables in the developmental system interact with one another and with particular variables. One can discover  $K$  by observing the variable’s tendency to

level off and approach some stable value (but see Abraham, forthcoming: attractors do not necessarily correspond with stable values).

#### DEFINING PARAMETER NAMES

An interesting feature of Excel is that it allows you to define names for specific parameters, or names for specific ranges. For instance, suppose you build a model in which you need two parameters, one that you wish to call growth rate of motivation, abbreviated as `rate_motivation` and one that you wish to call carrying capacity of motivation, abbreviated as `carcap_motivation`. With these named parameters, you would like to write equations like this  $= A2 + A2 * \text{rate\_motivation} * (1 - A2 / \text{car\_motivation})$ .

Assuming that the D and E columns are still empty, click on cell D1 and write

`rate_motivation`

do *not* use the `=` sign).

Click on cell D2 and write:

`carcap_motivation`

again without the equals sign.

Now select the range covering the cells D1–E2, which is a range consisting of four cells, two of which contain the names of the parameters and two of which are still empty. With these four cells selected, click on the menu item Formulas on top of the Excel sheet. A new menu bar will appear and in the middle you will find a group called Name Manager. Click on the option Create from Selection. A small pop-up window will appear, containing four options, and the option Left Column will be marked. That is to say, Excel assumes that you want to create names from values that appear in the left column of the selected range, which is indeed the case. Click OK.

Now select cell E1, which is the cell next to the name `rate_motivation`. Look at the name box, which is an editing box on the left of the worksheet, right under the menu bar. In the name box, the word `rate_motivation` should appear, indicating that cell E1 now has a new name, which is `rate_motivation`.

In cell E1, which now bears the name `rate_motivation`, type the value 0.1.

Go to an empty cell, for instance cell F1 and type `= rate_motivation`

You will see that cell F1 will show the result of this extremely simple equation, which is 0.1, the value of the cell `rate_motivation`.

If you change the value of the cell now named `rate_motivation` into 0.2, you will see that F1 shows the value 0.2.



In cell F1, now type  
`= rate_motivation*2`  
 and cell F1 will immediately show the result, which is 0.4. You can now use these newly defined range names into any equation you wish. You can define as many range names as you need. Working with defined range names will make your equations considerably more transparent to read, and changes to the values of the parameters will immediately affect the results of the dynamic models that refer to these parameters.

From a system-theoretical point of view, however,  $K$  typically represents *the more or less stable properties of the entire developmental system as it relates to a particular variable in the system* (e.g., reading comprehension, addition and subtraction skills, etc.). In developmental psychology, a person's genetic endowment has typically been seen as one of the constant properties of the developmental system that characterizes this person (although the way genes come to expression and eventually inhibit each other's expression is a very dynamical and complex process). Hence, a particular  $K$  factor, which applies for instance to musical or verbal skill, can be seen as a combination of more or less stable environmental factors and genetic factors. For all individuals for whom the general environmental factors are more or less the same, variability in the  $K$  factor typically corresponds with variability in the underlying genetic endowment  $G$ . In this way, the genetics can thus be entered into the dynamic growth equation, for example, an equation that describes the growth of reading comprehension, addition skill, musical skill, or a child's coordination and dissociation skills resulting after a history of abuse and neglect (Mascolo, Fischer, Van Geert, & Steenbeek, in press).

Since we are going to use this basic growth model as a component for building models of connected growers, the best thing we can do to model this in Excel is to first define a number of variable names, which you can do via the *Formulas/Create name from selection* menu option. Since this will be our first grower, define the variables `rate_1`, `car_1`, `ini_1`, corresponding with the rate parameter, the carrying capacity, and the initial value of the first grower. In cell A1 type

$$= ini_1$$

and in cell A2 type

$$= A1 + A1 * rate_1 * (1 - A1/car_1)$$

With this logistic growth equation, we can nicely show the difference between the notion of a flow and the notion of a map (see Abraham, forthcoming). In fact, in Excel all models are maps strictly speaking, because all processes take place in

the form of *discrete steps*. But we can approximate the notion of the flow by setting the values of the change parameters to very small numbers, and setting the cells to very small time steps. If the time steps are very small, we need a long sequence of time steps to model a sufficient time duration, for example, in the form of a column which is several hundreds of cells long. If we set the value of the change parameters to relatively high values, for example, between 1 and 3, we can experiment with regular and chaotic switches between the values, which are typical of the so-called logistic map (you can find an example in one of the Excel files that can be downloaded as associated with materials from [www.paulvangeert.nl/articles\\_appendices.htm](http://www.paulvangeert.nl/articles_appendices.htm).)

#### Step Four: Extending the Logistic Equation to Logistic Networks

In a developmental system, some of the variables are tightly connected. For instance, in most children, the connection between a child's addition skill and his subtraction skill is considerably closer than the connection between his addition skill and sports or musical skills, to name just two examples. Note that this type of connection should not be confused with the type of connections we are used to in the behavioral sciences, which usually take the form of an association between variables, often expressed in the form of a correlation, which is based on differences between individuals in a big sample. The connections we are focusing on here are *dynamic* connections. An example of a dynamic connection is when one variable facilitates or hampers the growth of another variable. For instance, in one particular child, insights achieved during the learning of addition might help this child in the learning of subtraction. In this example, addition skill has a positive, causal influence on the growth of subtraction skill. In another child, who has difficulty learning and needs to invest considerable time in practicing addition and subtraction skills, we might find a direct connection between addition skill and the skill of playing soccer. This child's soccer skills are based on his motivation to spend much time and energy practicing soccer, and this time and energy competes with the time and energy that the child invests in practicing addition and subtraction. This second child's dynamic network of growth relationships is thus different from the network of relationships between growers that is characteristic of the first child. This refers to an important property of dynamic modeling that we have discussed before, namely that the models must refer to individual growth trajectories, and that such individual growth trajectories may be highly idiosyncratic. The main question is: if a developmental system is in fact a network of dynamic connections, how can we model it by means of the logistic equation? The answer is that we can do so by means of the *extended logistic equation*.

Let us, for didactic reasons, begin with a very simple developmental network, namely one that consists of just two

variables. One variable corresponds with a particular math skill, such as addition, and the other variable corresponds with domain-specific motivation, that is, the child's interest in math and in particular in addition. In Excel, we define the parameters  $rate\_m$ ,  $car\_m$  and  $ini\_m$ , to refer to the growth rate and carrying capacity for an initial state of motivation for addition. Use the B column to set up a model of the growth of motivation. In order to turn this into a network, we must specify the relationships between these variables, that is, specify how one variable affects the growth of another. We shall assume that the level of motivation, that is, how motivated the child is to do this addition assignment, positively affects the growth of the addition skill. We can express this relationship in the form of a simple exponential term, similar to the exponential growth model we started with. For instance, if  $M_t$  and  $A_t$  are the levels of motivation and addition skill at time  $t$ , the effect of motivation  $M$  on the change in addition skill  $A$  is expressed as

$$A_t * (\text{effect\_of\_M\_on\_A} * M_t).$$

In Excel, the values  $A$  and  $M$  refer to the preceding cells in the corresponding columns, and the parameter  $\text{effect\_of\_m\_on\_A}$  has been defined in advance. For instance, if addition appears in column A and motivation skill appears in column B of an Excel worksheet, and the initial levels of addition and motivation skill appear in cells A1 and B1 respectively, the second step in the sequence of effects of motivation on addition is written as

$$A1 * \text{effect\_of\_M\_on\_A} * B1.$$

Let us now assume that the growth of motivation is affected by the *growth of addition skill* (that is not by the level of addition skill, but by the change in addition skill, that is, the amount of progress or regress in addition skill a person makes over some amount of time). The growth of addition skill is the difference between the current level of addition and the preceding one (alternatively we can define the *relative* growth of addition skill as the difference between the current and preceding level divided by the current level). Hence, the first equation for the change in motivation is

$$M_t * (\text{effect\_of\_A\_on\_M} * (A_t - A_{t-1}))$$

and the second, where the growth of motivation is affected by the *relative* change in addition skill is as follows

$$= \dots M_t * (\text{effect\_of\_A\_on\_M} * (A_t - A_{t-1}) / A_t).$$

Now we shall bring all the pieces together. To begin with, we have a logistic equation that describes the logistic growth in motivation and addition skill. Secondly, for motivation as well as addition skill we have an equation that describes how addition affects motivation and the other way around. We shall now bring these two parts together into extended logistic

equations, one describing the growth of (domain-specific) motivation and another the growth in addition, which results in the logistic network model for these two growers:

$$A_{t+1} = A_t + A_t * \text{rate\_A} * (1 - A_t / \text{car\_A}) + A_t * \text{effect\_M\_on\_A} * M_t.$$

$$M_{t+1} = M_t + M_t * \text{rate\_M} * (1 - M_t / \text{car\_M}) + M_t * \text{effect\_A\_on\_M} * (A_{t+1} - A_t).$$

Suppose again that addition occurs in the A column and motivation in the B column. The initial values of addition and motivation occur in cells A1 and B1, respectively. However, note that the model must calculate the change in addition in order to compute the effect of motivation. For this reason, we must put the initial values of addition and motivation into two successive cells, namely A1 and A2 for addition and B1 and B2 for motivation. In cell A3, we can write the first growth equation:

$$= A_2 + A_2 * \text{rate\_A} * (1 - A_2 / \text{car\_A}) + A_2 * \text{effect\_M\_on\_A} * B_2,$$

and in cell B3, the first equation for the change in motivation:

$$= B_2 + B_2 * \text{rate\_M} * (1 - B_2 / \text{car\_M}) + B_2 * \text{effect\_A\_on\_M} * (A_2 - A_1).$$

This simple example of connected growers (see Figure 2) can be expanded very easily to networks of many growers, and the reader can find several examples in the Excel files added as web materials ([www.paulvangeert.nl/articles\\_appendices.htm](http://www.paulvangeert.nl/articles_appendices.htm)). The bullying model described by Abraham (forthcoming) is an example of a network consisting of two connected variables described by growth equations. Fischer and Bidell (2006) provide several other examples of connected growers in the context of the theory of dynamic skills.

If the reader wants to specify models of connected growers in Excel, I recommend splitting every grower up into a number of constituents, which simplifies the task of writing the equations referring to many connected growers (e.g., see the Excel file available from the web materials [www.paulvangeert.nl/articles\\_appendices.htm](http://www.paulvangeert.nl/articles_appendices.htm)). I have written macros in Excel that can model networks of many connected growers, and that take matrices of relationships as their input (see the Excel file available from the web materials). Networks of connected growers can also—and probably more easily—be modeled in the form of dedicated systems-dynamics software, which works with stock and flow components, such as the stock and flow options under Net Logo discussed by Abraham (forthcoming) or the freely available Insight Maker software that runs under a web browser, or the commercial packages *Stella* and *Powersim*.

An interesting feature of these network models is that they generate a number of qualitative properties that are

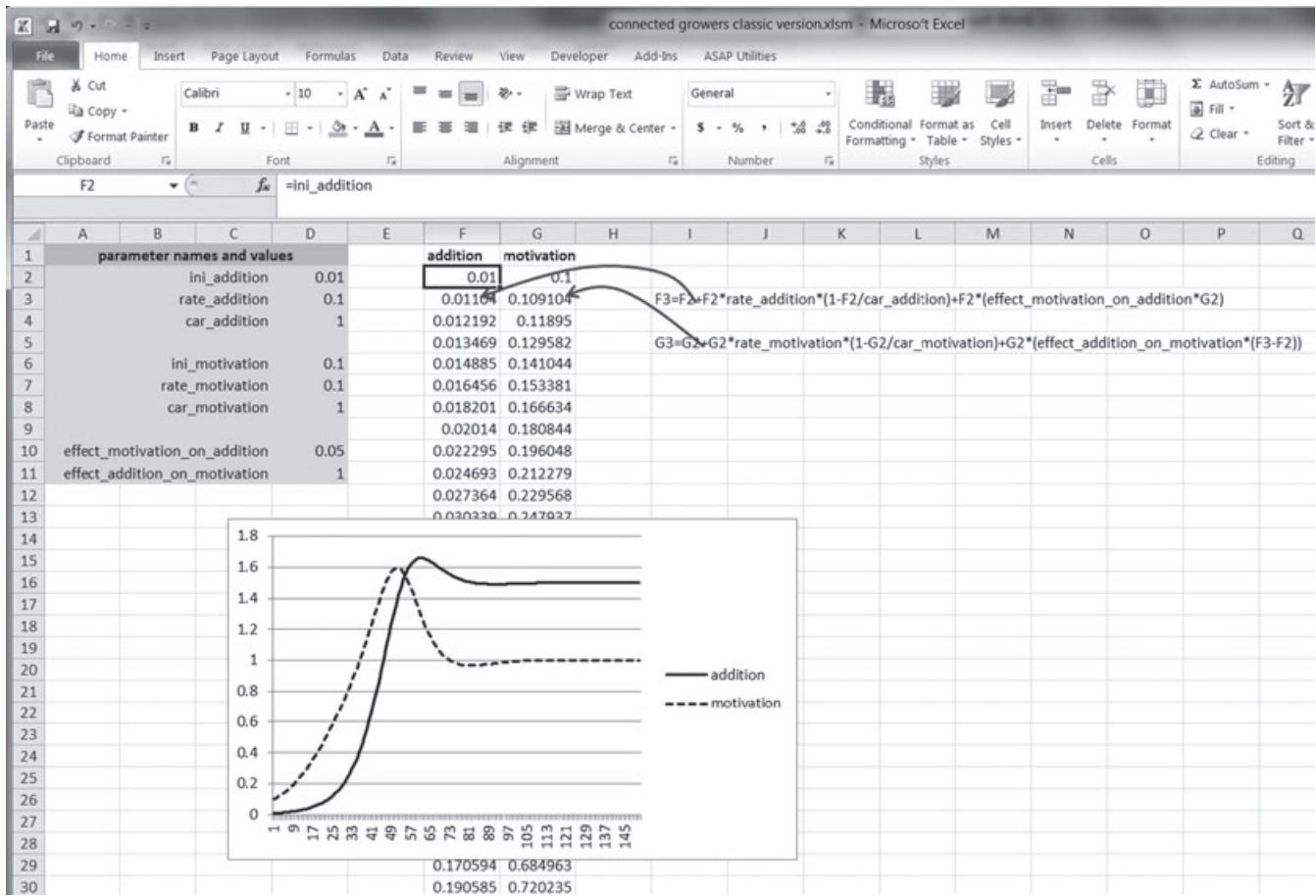


Fig. 2. Screen picture of the network of two connected growers, namely addition skill and its domain specific motivation. The nature of the connections between the two variables typically results in “overshoot” values and stabilization at the lower than the maximum level.

characteristic of developmental systems in general, such as the occurrence of stepwise changes, of overlapping waves, of temporary regression and of the emergence of stable states. Overlapping waves, for instance, typically emerge in networks where one variable represents a more primitive or less well-developed way of accomplishing something. Examples are the overlapping waves in mathematical strategies that differ in terms of cognitive complexity, or overlapping waves in forms of moral reasoning that represent different developmental levels. Another example is the use of one-word sentences, 2–3 word sentences, or 4+ word sentences, which represent different levels of grammatical complexity (Bassano & Van Geert, 2007). In such a network of variables, the less developed strategy or skill has a positive influence on the emergence of a strategy or skill of higher complexity, in that the more complex strategy or skill integrates the less complex one. Conversely, the more complex strategy has a negative effect on the less complex one, in that the more complex strategy can easily replace the less complex one, for instance because it is more versatile. The Excel file available from the web materials (*sentence length.xlm*) provides an example of a model that has been fitted to the empirical

data and provides the relevant Excel equations to model this and comparable networks giving rise to overlapping waves.

Temporary regressions typically emerge in models where variables are negatively affected by the *change* in other variables. For instance, the learning of subtraction can in principle temporarily suppress a child’s addition skill, but as soon as the subtraction skill is more or less consolidated, the negative effect disappears (e.g., Feldman & Benjamin, 2004). Such negative effects are likely to occur if a particular process of learning recruits considerable attentional resources, or requires restructuring of problem-solving strategies. That is, children can be temporarily confused by the learning of a new skill, which has a temporary negative effect on the old associated skills, such as addition and subtraction skills (for an example of temporary regressions and stepwise growth of cognitive skills, see Fischer & Bidell, 2006; see also Figure 3, modeling coordination and dissociation skills in children suffering from serious traumatic abuse).

Growth models have been explained in the context of language development and cognitive growth, and examples are given in a variety of articles and chapters (Van Geert, 1991,

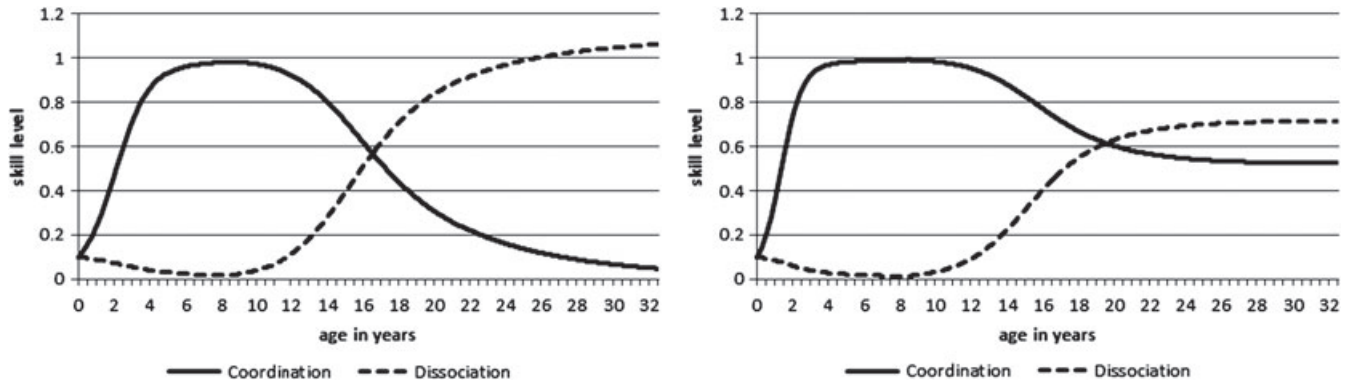


Fig. 3. The simulation of two trajectories of coordination and dissociation skills in a child who suffers from serious traumatic abuse from the age of about 8 years on (the model’s step size is one month). The first figure shows a trajectory of dissociation following trauma at around the age of 8 years which stabilizes well into adulthood and becomes the sole way of coping for the person in question. The other figure shows a trajectory after a similar trauma that shows a pattern of stabilization between coordination and dissociation, with coordination still active, but with dissociation as the dominant mode of adaptation.

2003; Van Geert & Fischer, 2009; Van Geert & Steenbeek, 2005a, 2005b, 2013). In the remainder of this article, I will present an example of a developmental question—why do girls like pink—and of the steps that can be taken to transform a developmental question into a conceptual model of change, and the conceptual model of change into a mathematical model that allows the maker to simulate developmental scenarios under various assumptions.

### WHY DO GIRLS LIKE PINK?

Just recently I was called by a local radio station and they asked me if I could answer a question from a young listener, who wanted to know why girls like pink. Do girls like pink? Looking for data answering this question, I also asked myself whether the alleged preference of girls for pink changes across the lifespan. Do baby girls prefer pink, and do adult girls also prefer pink? But what does this preference mean? A particular girl, maybe the one who called the radio station, might prefer pink dresses and blue purses, and prefer fair hair over pink hair. And at times she might prefer a blue dress over a pink one. All these phenomena have to do with variation over time, and variation over time is the domain of dynamic systems and dynamic systems modeling.

While preparing my answer to this question I found an interesting article by LoBue and DeLoache (2011), which had cross-sectionally investigated the preference of boys and girls for pink versus blue between the ages of 1 and 5 years. The age curve of proportions of pink over blue choices shows an interesting nonlinear pattern, and we shall ask ourselves whether we can build a dynamic model explaining it.

Figure 4 shows cross-sectional data, aggregated over many different individuals. In principle, dynamic models describe actual processes, for instance, the change of preference for

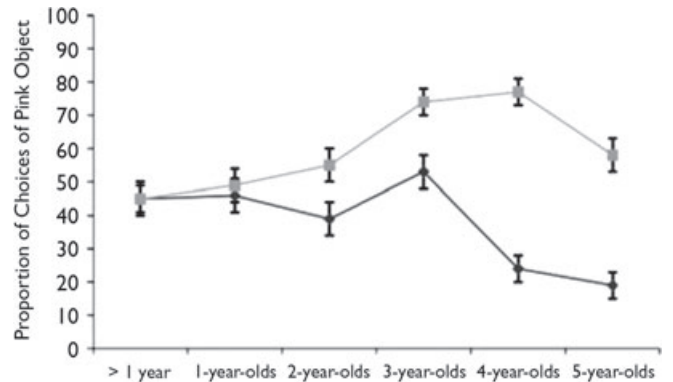


Fig. 4. Emergence of preference for pink in boys and girls between the ages of 15 years (after LoBue & DeLoache, 2011).

pink during early childhood. Since the change of preference is a process that takes place in individuals, the dynamic model must generate *individual trajectories*. Such trajectories need not be similar to the trajectory specified by the cross-sectional data, and neither should they be similar to the average of the individual trajectories (for discussion, see Molenaar & Campbell, 2009, on the issue of non-ergodicity in behavioral developmental data; see also Rose, Rouhani, & Fischer, 2013). If the dynamic model is used to understand cross-sectional data averaging over groups, it must first generate a great many individual trajectories, and then calculate averages for each step in this collection of trajectories. This is different from the way a statistical model might be used to explain the data. The statistical model estimates the *average value* of each point in the cross-sectional curve by expressing it as a function of the independent variable, which in this particular case is age. However, a good dynamic model must not only be able to explain the cross-sectional data based on generating many individual trajectories, but also the important

qualitative properties of such individual trajectories. A model of individual trajectories will be able to explain cross-sectional data, but a model of cross-sectional data—or, for that matter, longitudinal data with relatively few measurement points—cannot explain individual data, unless the process is ergodic (the ergodic process is one in which the statistical structure of inter-individual variability is similar to the structure of intra-individual variability, which is hardly ever the case for behavioral data).

The data show that at the age of approximately 1 year on average girls and boys do not differ as regard their preference for pink. That is to say, there is no evidence for an early gender-specific preference, which is likely to be due to biological or genetic factors (see also Franklin, Bevis, Ling, & Hurlbert, 2010). However, the lack of an early preference does not imply that there exists no genetically determined color preference between men and women. Such genetic preferences might become manifest as children grow older and become adults (there is plenty of research showing that for a broad range of psychological properties, the heritability component is increasing with increasing age; Davis, Haworth, & Plomin, 2009; Haworth et al., 2010; Johnson, 2010). With regard to color preference, research has shown that, although adults in general prefer blue over other colors, women tend to prefer more reddish shades of a particular color than men do (Hurlbert & Ling, 2007).

### A Dynamic Model Explaining the Emergence of Pink Preference as a Result of Social Influences From High-Power Social Models

LoBue and DeLoache (2011) claimed that the emergence of gender-specific preference is primarily due to processes of modeling and imitation. That is, children tend to imitate the choices of same-gender models or peers, and tend to oppose (in a sense “counter-imitate”) the choices of different-gender models or peers. In the real world, children are influenced by different ages and social status, including social models that feature in gender-specific commercials, for example, for toys. It is highly likely that as children grow older, they become more sensitive to such high-power social models, including those in commercials.

Let us assume that the color preference displayed in TV commercials, and by other high-power examples, such as parents or adults, serves as an *attractor* for the color preference of young children (for the notion of attractor, see Abraham, forthcoming). On the basis of the study by LoBue and DeLoache (2011) we shall assume that on average the early preference for pink in boys and girls is 50% (i.e., no selective preference for pink). However, what is the preference value of the attractor, which boys and girls are supposed to infer from the high-power examples? What is the color preference represented in commercials, or shown by peers, and parents?

Is it a 60% preference of pink for girls, or maybe 70%, 80%, ...? The answer is that we don't know, but we can of course make a reasonable guess, for instance that the modeled preference is 70%. Once the model is built, we can easily change this parameter to check how sensitive the model is to the exact value of this parameter. For the dynamic model, what matters is not so much the correctness of the guess, but rather the correctness of the model it provides for the trajectory that leads girls to move towards the socially imposed female preference, and that leads boys to move away from the socially imposed female preference. If we assume that boys tend to oppose the preference of girls, we can represent the socially imposed preference of boys as the inverse of the socially imposed preference of girls, which in this particular case is  $100\% - 70\% = 30\%$ . We already assumed that children tend to become increasingly aware of the presence of high-power social models, such as their parents, older children, or children in TV commercials. That is, social imitation of high-power models becomes increasingly important as children grow older. This means that we need to develop a dynamic model of the *growing sensitivity towards gender-specific modeling*.

Let us assume that the sensitivity to social models (regarding color preference) is minimum at birth, and that it reaches a maximum around the age of about 5 years (this is just a guess, which might be replaced by any other guesses based on empirical research; what matters here is a demonstration of the underlying principles of dynamic modeling). The point is that we have no clues as to how this sensitivity actually increases. Let us assume that the increasing sensitivity can be explained by the logistic growth function, which is general enough to capture the underlying mechanisms. Assuming that cell A1 contains the initial value of the *sensitivity*, and that you have defined a cell which is named *rate1*, and that the maximum sensitivity equals 1, cell A1 should contain

$$= 0.01$$

and cell A2 should contain the following equation:

$$= A1 + A1 * rate1 * (1 - A1)$$

(define the value of the parameter *rate1* as 0.16).

Copy cell A2 to a range from A3 to A50 to obtain a column that covers a duration of 5 years with a step size of 1/10 of a year.

The next thing is to actually model the growth of preference in girls, which we will do in the B column. In cell B1, write the initial value for the preference for pink, which, according to the empirical data, is 0.5. We shall also assume that the growth of preference is governed by the logistic growth equation, which seems like a fair assumption in view of the general nature of the underlying logistic model. First, define a variable called *r\_p\_G* (which stands for rate of preference in girls). Next, define a parameter called *P\_G*, which stands for the female preference

for pink as demonstrated by high-power models, advertisements, television programs, and so forth (recall that I suggested a value of 0.7, i.e., 70% for this parameter). Then, in cell B2, write the equation for the next level of preference in girls

$$= B1 + B1 * r_{p\_G} * (P\_G - B1).$$

This equation assumes that from the start on, children are fully sensitive to preference modeling by others. However, the actual growth of preference is a function of the growth rate *and* of the sensitivity for social models. That is to say, in the model, the growth of preference must be coupled to the growth of sensitivity for social influences, which can be accomplished as follows. At time step 2, we can find the corresponding level of sensitivity for social models in cell A2. Hence, we can specify sensitivity-dependent growth of preference as follows

$$= B1 + B1 * r_{p\_G} * (p\_G - B1) * A2.$$

If we copy this equation to cells B3–B51, we have a dynamic model for the growth of preference for pink from birth until the age of 5 depending on increasing sensitivity for social models. A line plot of these 51 cells shows an S-shaped pattern of growth of preference for pink, stabilizing at a preference of 0.7. Note that this pattern applies to a single girl, that is, to an individual trajectory. If all trajectories are similar, a single girl represents all girls, but in practice, individuals differ in the value of the growth parameters and initial states. It is very likely that they also differ in their perception of the attractor state. Some girls might overestimate the effect of the social and media models, whereas others might underestimate it. Thus, in order to simulate a population, let's say of 10 individuals, you can write 10 double columns (one for the sensitivity, the other for the preference) based on different parameter values that you think are representative for the differences in the population under study. To define the parameter values, you can use the equation for random numbers drawn from normal distributions, assuming that the parameters are indeed normally distributed. You can find the details of this model in the associated Excel file, available from the web materials ([www.paulvangeert.nl/articles\\_appendices.htm](http://www.paulvangeert.nl/articles_appendices.htm); see the file *preference for pink.xlsm*).

Empirically speaking, preferences cannot be measured directly, and have to be inferred on the basis of choices made by a person with a particular preference. The preference of 0.7 means that if you make many color choices, approximately 70% of those choices will be for pink. Choosing a particular color, which in this case means either pink or some other color, is easy to simulate. We have already seen that Excel (and any other system dynamics simulation program) contains an equation that calculates a random number between 0 and 1, namely `Rand ()`, with all values occurring with equal probability. If we define a choice for pink as 1, and the choice for another color as 0, we can simulate a choice for pink by means of the

following equation, which we enter in cell C2 (recall that the corresponding color preference appears in cell B2:

$$= \text{if} (\text{Rand} () < B2, 1, 0).$$

If you recalculate this many times, you will see that the proportion of 1's over 0's is approximately similar to the preference value, which you find in cell B2. Copy this equation to cells C3–C51, which means that for each time step (and there are 10 such steps in a year, according to our model) you obtain a random choice for pink or for another color. By calculating the average over the 10 steps that correspond with a particular year, you obtain 5 values, and each value provides an estimation of the preference for pink, for a particular girl, during a particular year. If you recalculate this model several times (you do so by pressing the F9 button), you will observe rather significant differences between the estimated preferences, but this is of course due to the stochastic nature of the actual choice process. It is easy to see that if you model 10 girls and 10 boys, the averages of all their choices, will give you a better estimation of the *average* preferences during a year, and of the average differences between the boys and the girls.

In order to model the growth of the preference for pink in boys, you have to take into account that the preference for pink in the boys is the opposite of the preference for pink in the girls, namely  $1 - P\_G$ .

### A Dynamic Model Explaining the Emergence of Pink Preference as Imitating the Preferences of Peers

In the previous dynamic model, we assumed that the gender-specific preference for pink is represented by high-power social gender models such as older children, adults in real life, and in commercials. However, it is likely that gender-specific preference for pink emerges as a consequence of the fact that children tend to imitate the choices of same-gender peers, and tend to reject the choices of other-gender peers. Because children meet their peers on a regular basis, in daycare centers for instance, they can actually observe what other children are doing, and imitate their choices in a gender-specific way. The dynamic model that we shall construct will be able to demonstrate that gender preferences may spontaneously emerge through peer imitation, even if, at the beginning of the developmental processes, no such preferences actually exist in the peer group.

The model begins with a definition of two groups of children, namely a group of 10 boys and a group of 10 girls (the reader can construct a model with smaller or larger groups, but 10 is a good didactic compromise). For simplicity, we focus only on the effect of the behavior of others, and thus refrain from modeling the sensitivity aspect (once we have this model running, we could extend it with a component featuring the sensitivity aspect). We begin by giving the boys and girls the same initial preference for pink, 0.5.

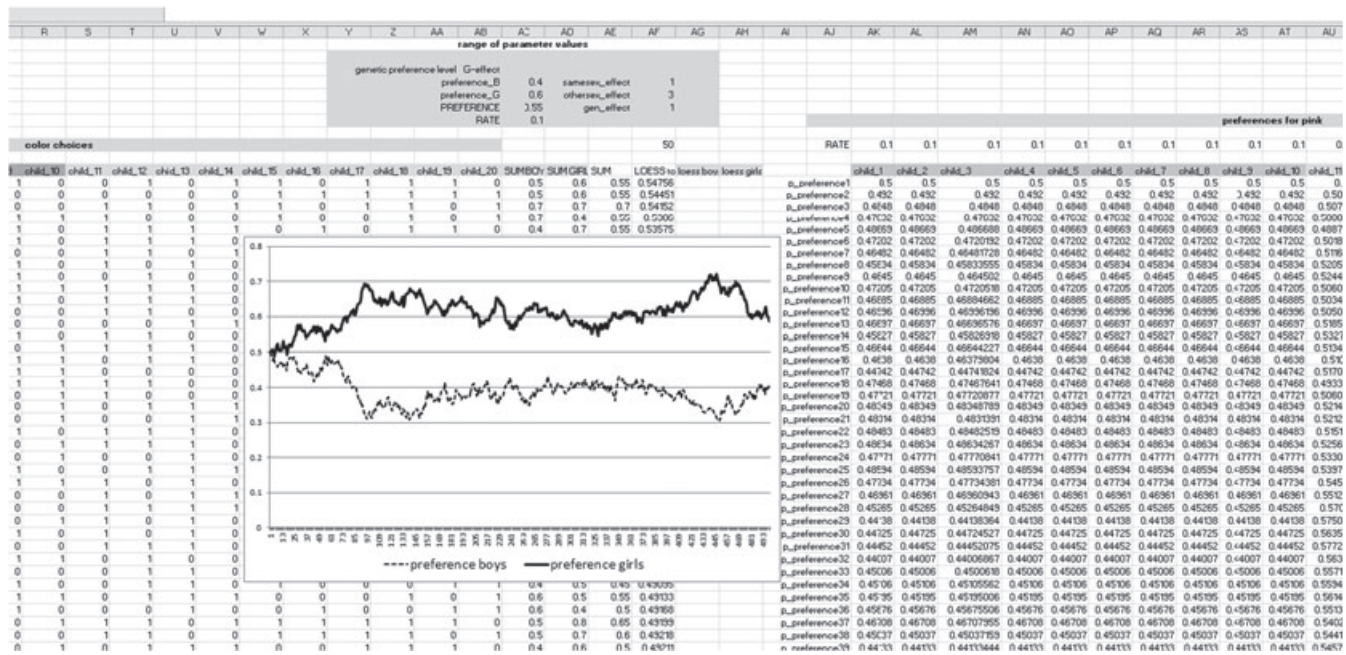


Fig. 5. A screen picture of the Excel model of the emergence of gender-specific preference for pink in boys and girls. The model simulates the preferences (right) and the associated choices (left) in 10 boys and 10 girls. Preference changes on the basis of the children’s perception of the average same-sex choice and the average other-sex choice. The model simulates the effect of the tendency to imitate the choice of same-sex peers and avoid the choice of other-sex peers. In this model, the weight of the same-sex effect, the other-sex effect, and the genetic effect can be varied.

In another range on the same Excel worksheet, covering the same 10 boys and 10 girls, we model their choices for pink or for another color during each time step, dependent on the corresponding preference (see Figure 5). All this is done in a way comparable to that described for the first model, which was based on the idea of increasing sensitivity to gender-specific examples. Finally, let us, for simplicity, assume that there are no innate differences in gender specific color choices (it is easy to add this assumption later in the model).

The model assumes that, in this particular group, the 10 boys and 10 girls are having enough social interactions to allow each child to observe the color choices occasionally made by his or her 19 peers, 9 of which are of the same gender and 10 of which are of a different gender. We also assume that these observations are in a sense averaged by the children and kept in the children’s memory. These assumptions are in line with recent attempts towards explaining infant learning as forms of statistical learning (see for instance Denison, Reed, & Xu, 2013; Kirkham, Slemmer, & Johnson, 2002; Saffran, 2001; Vlach & Johnson, 2013; Xu & Kushnir, 2013). Hence, we define two new columns, one containing the average number of pink choices made by the boys, the other the average number of choices for pink made by the girls, for each time step. In this model, the attractor value for pink in boys and girls at a particular time *t*, is defined by the current same-sex choice for pink and by the inverse of the other-sex choice for pink. The model, which can

be found in the web materials, clearly shows that a gender-specific preference for pink emerges as a result of processes of choice that started with no preference at all. In the simplest version of this model, there is no biological assumption about gender differences in sensitivity to particular colors. As a result, some simulations show an emergent preference for pink in girls, others show an emergent preference for pink in boys. Adding a genetic preference component to the model (see the Excel file Preference for pink.xlsmin in the web materials) reduces the probability that a “wrong” preference emerges.

### WHY MODELING AND HOW TO DO IT? SUMMARY AND CONCLUSIONS

Dynamic modeling is an important tool for helping researchers understand the nature of developmental learning trajectories over time. In order to build a model, the researcher must focus on the principles that govern the *change* in the variables that he or she is interested in. Dynamic models explain change by making explicit how a particular current state of the system—in the form of a collection of variables that applies to an individual person, dyad or group (not a statistical sample)—is transformed into the next state of the system. Dynamic models apply these principles of transformation in an iterative fashion (the output of the preceding step is the input for the next step in the process) and by doing so generate

explicit, numerical descriptions of temporal trajectories of the variables at interest.

Modeling is far more than number crunching and fitting of empirical data. In fact, it primarily deals with conceptual and theoretical issues. That is to say, to a considerable extent, modeling is a way of experimentally studying one's theoretical and conceptual ideas, i.e., it is a form of "theoretical experimental psychology" (Van Geert, 1994). It is "experimental" because, once you have a model, you can experiment with a variety of conditions and parameter values, in order to discover what consequences they have for the model outcomes. This is what modelers often call experiments *in silico*, as opposed to *in vivo* (see, e.g., Schumacher, Ballato, & Van Geert, forthcoming).

Spreadsheets, in particular the widely used Excel software, provide easily accessible and insightful ways for building dynamic models. Although they have certain disadvantages—they can approximate real flows (see Abraham, forthcoming) only to a limited extent—they also have considerable advantages. One is their wide availability and the fact that for most users they will require very little initial learning. The second is that they represent the time steps in the model in the form of a succession of cells on a worksheet, thus visualizing the temporal trajectory in a literal way. They make it easy to combine mathematical equations with random numbers and if-then statements.

Dynamic models must be based on conceptually adequate principles of change that apply to learning and development. A highly generalizable principle of change is the principle of logistic growth, which is based on the assumption that the change in a particular developmental or learning variable is, on the one hand, a function of the level already attained (the current state), and on the other hand, a function of limited resources. These resources comprise everything that either negatively, positively or conditionally affects the change of a particular variable. They can be distinguished into resources that are relatively constant on the time scale at which the change in the variable at interest occurs, and resources that are typically variable on this time scale. The constant resources constitute the carrying capacity parameter of the logistic growth model, including for instance genetic factors. The changeable resources—which are often reciprocally coupled with the variable the growth of which we aim to model—typically constitute the variable support and competition component of the logistic growth equation.

The logistic equation, including the relationships of support or competition with other variables, can be used to specify dynamic network models. Network models provide powerful tools for modeling a wide variety of developmental and learning trajectories. Network models can range from two to any number of connected variables.

By means of a specific example—the emergence of color preference in children—we have demonstrated how models

of concrete developmental or learning phenomena may be constructed. The primary aim of building a dynamic model is to obtain a better understanding of the principles explaining the form of the developmental trajectories. In order to achieve this aim, model builders must start from a conceptual or theoretical model of the phenomenon at issue, and ask themselves how these conceptual or theoretical principles can be transformed into mathematical equations (Brown, Sokal, & Friedman, 2013). Second, model builders must try various scenarios to see which principles of change are connected with which forms of developmental change. I have given an example of this procedure by separately studying the effect of imitating high-power social models and the effect of imitating peers. By starting from the assumption that there is no genetically determined color preference at the beginning of development, we were actually exploring to what extent assumptions regarding innateness are necessary to explain gender-specific color preferences. Third, model builders must primarily focus on obtaining a fit between the model results and important *qualitative* features of the range of empirical phenomena that they wish to understand. For instance, an important qualitative feature of the emergence of gender-related color preference is that it develops from an initial state in which gender differences are absent. Another example of an important qualitative feature, for instance in cognitive development, is the occurrence of temporary regressions in performance preceding a leap towards a higher level of skill (e.g., Feldman & Benjamin, 2004; Fischer & Bidell, 2006). Once a dynamic model provides a conceptually adequate explanation of the major qualitative features characterizing a broad range of empirical phenomena, it can also be used to try to achieve a fit with a specific data set.

#### NOTE

- 1 Riding a bicycle is typically a multidimensional skill, involving the coordination of many perceptual and motor components. However, in this multidimensional space, we can define a distance between a "novice" region (those who cannot ride a bicycle) and an "expert" region (those who can ride a bicycle). The change in the skill can then be defined in terms of this one-dimensional distance variable.

#### REFERENCES

- Abraham, R. (forthcoming). Agent-based modeling of growth processes. *Mind, Brain, and Education*.
- Bassano, D., & Van Geert, P. (2007). Modeling continuity and discontinuity in utterance length: A quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science*, *10*, 588–612.



- Brown, N. L., Sokal, A. D., & Friedman, H. L. (2013). The complex dynamics of wishful thinking: The critical positivity ratio. *American Psychologist*, 68, 801–813.
- Davis, O. S. P., Haworth, C. M., & Plomin, R. (2009). Dramatic increase in heritability of cognitive development from early to middle childhood: An 8-year longitudinal study of 8,700 pairs of twins. *Psychological Science*, 20, 1301–1308.
- Deci, E. L., & Ryan, R. M. (2009). Self-determination theory: A consideration of human motivational universals. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (pp. 441–456). Cambridge, England: Cambridge University Press.
- Dehaene, S. (2007). A few steps toward a science of mental life. *Mind, Brain, and Education*, 1, 28–47.
- Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, 49, 243–249.
- Devlin, K. (2013). *What exactly is multiplication?* Mathematical Association of America website. Retrieved from <http://archive.is/qCkK>
- Feldman, D. H., & Benjamin, A. C. (2004). Going backward to go forward: The critical role of regressive movement in cognitive development. *Journal of Cognition and Development*, 5, 97–102.
- Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action and thought. In W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development. Handbook of child psychology* (Vol. 1, 6th ed., pp. 313–399). Hoboken, NJ: Wiley.
- Franklin, A., Bevis, L., Ling, Y., & Hurlbert, A. (2010). Biological components of colour preference in infancy. *Developmental Science*, 13, 346–354.
- Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J. C., van Beijsterveldt, C. E. M., ... Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, 15, 1112–1120.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41, 111–127.
- Hillman, E. M. C. (2007). Optical brain imaging in vivo: Techniques and applications from animal to man. *Journal of Biomedical Optics*, 12(5), 051402.
- Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology*, 17, R623–R625.
- Johnson, W. (2010). Understanding the genetics of intelligence: Can height help? Can corn oil? *Current Directions in Psychological Science*, 19, 177–182.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35–B42.
- LoBue, V., & DeLoache, J. S. (2011). Pretty in pink: The early development of gender-stereotyped colour preferences. *British Journal of Developmental Psychology*, 29, 656–667.
- Mascolo, M., Fischer, K., Van Geert, P., & Steenbeek, H. (in press). What can dynamic systems models of development offer to the study of developmental psychopathology? In D. Cicchetti (Ed.), *Developmental psychopathology* (3rd ed.). New York, NY: Wiley.
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18, 112–117.
- Rose, L. T., Rouhani, P., & Fischer, K. W. (2013). The science of the individual. *Mind, Brain, and Education*, 7, 152–158.
- Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, 81, 149–169.
- Schumacher, N., Ballato, L., & Van Geert, P. (forthcoming). Using an agent-based model to simulate the development of risk behaviors during adolescence. *Journal of Artificial Societies and Social Simulation*.
- Steenbeek, H. W., & Van Geert, P. L. (2007). A theory and dynamic model of dyadic interaction: Concerns, appraisals, and contagiousness in a developmental context. *Developmental Review*, 27(1), 1–40.
- Steenbeek, H., & Van Geert, P. (2008). An empirical validation of a dynamic systems model of interaction: Do children of different sociometric statuses differ in their dyadic play? *Developmental Science*, 11, 253–281.
- Steenbeek, H., & Van Geert, P. (2013). The emergence of learning-teaching trajectories in education: A complex dynamic systems approach. *Nonlinear Dynamics, Psychology, and Life Sciences*, 17, 233–267.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.
- Van Dijk, M., Van Geert, P., Korecky-Kröll, K., Maillochon, I., Laaha, S., Dressler, W. U., & Bassano, D. (2013). Dynamic adaptation in child–adult language interaction. *Language Learning*, 63, 243–270.
- Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, 98, 3–53.
- Van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos*. New York, NY: Harvester.
- Van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky and beyond. *Psychological Review*, 5(4), 634–677.
- Van Geert, P. (2003). Dynamic systems approaches and modeling of developmental processes. In J. Valsiner & K. J. Conolly (Eds.), *Handbook of developmental psychology* (pp. 640–672). London, England: Sage.
- Van Geert, P., Den Hartigh, R., Steenbeek, H., & Van Dijk, M. (submitted). *The development of expertise and exceptional performance: A dynamic network model*. Manuscript submitted for publication.
- Van Geert, P. L. C., & Fischer, K. W. (2009). Dynamic systems and the quest for individual-based models of change and development. In J. P. Spencer, M. S. C. Thomas, & J. McClelland (Eds.), *Toward a new grand theory of development? Connectionism and dynamic systems theory reconsidered* (pp. 313–336). Oxford, England: Oxford University Press.
- Van Geert, P., & Steenbeek, H. (2005a). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. *Developmental Review*, 25, 408–442.
- Van Geert, P., & Steenbeek, H. (2005b). The dynamics of scaffolding. *New Ideas in Psychology*, 23(3), 115–128.
- Van Geert, P. L. C., & Steenbeek, H. W. (2013). (P)(re) presentations: What are they and how do they develop? Commentary on “Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate” by J. Allen & M. Bickhard. *Cognitive Development*, 28, 138–143.
- Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants’ cross-situational statistical learning. *Cognition*, 127, 375–382.
- Walker, J. T. (1996). *The psychology of learning*. Upper Saddle River, NJ: Prentice Hall.
- Weisstein, E. W. (1999). *CRC concise encyclopedia of mathematics*. Boca Raton, FL: Chapman & Hall/CRC.
- Xu, F., & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Directions in Psychological Science*, 22(1), 28–32.