

Variations on Sudoku

When we build an EM artefact relating to a particular domain, there is typically no specific functionality in mind. A well-constructed artefact will admit many extensions and new interpretations. This can be illustrated with reference to work that has been done by several students (including Karl King, George Efstathiou, Antony Harfield and Eric Chan) on the theme of the Sudoku puzzle. (Basic knowledge of this by now well-known puzzle is assumed!) Several possible directions in which their models can be extended and applied will be briefly discussed. (For a review of some of these you can consult EM paper 100: Lifelong Learning, Empirical Modelling and the Promises of Constructivism.)

A basic EM Sudoku model was first constructed by Karl King (see [sudokuKing2006](#)). In its original form, Karl's model gave some support to the human solver - for instance, for each blank square in the grid, listing all the digits that did not appear in the same row, column or region. Later versions of this model gave further information about what can be inferred in a given situation through applying the most basic rules. For instance, the model would indicate when there was only one possible digit that could be placed in a square taking account of the basic constraint mentioned above, or identifying where there was only one position in any particular row, column or region for a specific digit.

The Sudoku puzzle is linked to the problem of matching in a bipartite graph, to which Phillip Hall's "Marriage Theorem" applies. George Efstathiou, who added a definitive notation for specifying combinatorial graphs to EDEN, used the family of matching problems associated with a Sudoku puzzle to illustrate the use of his notation. For each Sudoku position, there is a matching problem on a bipartite graph whose vertices comprise the 9 squares in a particular row, column or region and the nine digits that can be allocated to them. These 27 matching problems are of course interdependent. (See [sudokuEfstathiou2006](#) and Efstathiou's MSc thesis Chapter 6, p90.)

A further extension of [sudokuKing2006](#) was implemented by Antony Harfield in the form of [sudokucolourHarfield2007](#). This model associates a colour with each digit from 1 to 9, and blends these colours to represent a subset of the numbers 1 to 9. This makes it possible to colour the squares of the grid in such a way that each blank square has the blend of colours associated with digits that are not already represented in the associated row, column or region. In its original form, the colours assigned to digits could only be changed by redefinition via the EDEN interface. A subsequent extension enabled the RGB components of the nine basic colours to be manipulated by sliders. (This made use of the gel notation.) In this form of the model, changing the colour assignment makes it relatively easy to identify whether there is but one place in a given row, column or region where a particular digit can be placed.

A program derived from [sudokucolourHarfield2007](#) (with the basic functionality afforded by the GUI in the EDEN model, but implemented in FLASH) can be found at warwick.ac.uk/go/sudoku/. This is intended to be a public demonstration in some way representative of what Empirical Modelling involves, but of course - as a conventional program - it is in some ways entirely misleading. It has very recently become possible to present a variant of [sudokucolourHarfield2007](#) using Web-EDEN.

One of the motivations for devising the colour variant of the Sudoku model was simply to explore colour display in graphics. The choice of basic colours to associate with digits and the strategy for blending them is open for experiment. One interesting possible extension might involve overlaying the colours on the International Commission on Illumination (CIE) diagram developed through empirical studies of colour vision in 1931. In this space, particular hues are represented by specific points, and a blend of colours by points within their convex hull. There is a curious possibility here of developing a practical application for [linesBeynon1988](#) in which issues such as determining when two 'independent' blends yield the same colour.

Another aspect of the study of Sudoku is concerned with solving strategies and associated cognitive issues. There are some interesting relevant reflections in an article by Brian Hayes in the American Scientist: see <http://www.americanscientist.org/template/AssetDetail/assetid/48550>. This can be more broadly set within the context of creating and solving Sudoku puzzles. Under "creating", we might want to explore what justifies calling a Sudoku puzzle "hard", "medium" or "easy". Under "solving", there is a "pupil" perspective, concerned with getting the answer, and a deeper educational perspective concerned with issues such as "teaching problem-solving" or investigating the nature of the learning activity involved. The idea that an EM model can provide unifying support to the perspectives of the developer, learner and teacher is well-represented in EM papers on the theme of

educational technology. (See Sudoku.ppt for list of potential observables that relate to these three perspectives.)

From an EM perspective, an interesting question relating to these agendas is how we understand "dependency" in this context: after all, if a puzzle is well-posed, the content of every square is determined by the given digits, and can thus be regarded as defined by a dependency. In contrast, the dependencies that can reasonably (or perhaps only arguably) be said to be "apprehended in experience" are those that correspond to simple applications of rules. One possible direction for exploration (applicable only to hard puzzles - if at all!) concerns collaborative solving strategies. One suggestion that has been briefly studied by Eric Chan is that getting individual solvers to record dependencies of the form "if there's a 1 in this square, there's a 2 in that square etc etc", and to collate these in a corporate solving exercise. This is the kind of activity for which an EM model is well-suited, since it is possible to frame definitions that express such possible deductions without specifying whether the premise is true. (See the file CollabSolve.html for details.)

Studying the cognitive aspects of solving from an educational perspective is another theme. We understand relatively little about what possible advantages there might be in using the colour model for solution for instance. We might suppose that adding this experiential aspect might help a human solver who is not given to meticulous logical argument, but it is unclear whether in fact the "intuitive approach" to solution is really a sloppy form of guessing that inappropriately discourages reasoning. We have some empirical data from two visits by parties of 20 schoolchildren that might be useful in this connection. Some progress has been made on using EM to gather data for an empirical study of what strategies the schoolchildren actually used in tackling puzzles with sudokuKing2006 and sudokucolourHarfield2007 (with one or two simple extensions added by Beynon in 2008). There is a connection here with tools for recording and replaying interactions with EM models that have been investigated over many years.

When we look at possible extensions and adaptations of the existing EM Sudoku models in this way, we become aware of technical deficiencies in these models. For instance, Russ has proposed that removing the digits from the colour Sudoku model might lead to an interesting variant. To support this, it would be good to enlarge the squares of the Sudoku grid but this is surprisingly inconvenient because of the internal representation chosen by King. A quite independent approach to constructing Sudoku puzzles - and generalisations - can be found in the *Is Empirical Modelling puzzling?* contribution to WEB-EM-02.