Fingers or rods? An experiment in using AI to recognise placements of Cuisenaire rods and aurally endorse children's mathematical re-worldings

Michael Rumbelow

University of Bristol, UK

A prototype app was developed that uses AI to recognise placements of Cuisenaire rods on a table via a webcam and generates pre-programmed utterances, such as the lengths of the rods. A theoretical framework is proposed which brings together Wittgenstein's concept of a language-game, consisting of language and its interwoven actions, with Froebel's Kindergarten pedagogy of children creating 'worlds' by interweaving block play and story-telling. In this framework Froebelian 'worlding' can be understood as enacting language-games around block play, and the rules of a language-game, or 'worlding', around mathematical block placements can be endorsed by the app's utterances. This framework was used to design an experiment carried out with primary school children. I reflect on the results, focusing on a video-clip of a 6-year-old's interactions with the app, and tentatively identify evidence of a mathematical 're-worlding', transforming rods from counters into lengths.

Keywords: artificial intelligence; kindergarten; manipulative materials

Introduction

In this paper I describe development of a prototype AI app which detects children's arrangements of Cuisenaire rods on a tabletop via a webcam and 'speaks' the lengths of the rods. Secondly, I outline a theoretical framework for an experiment design, seeing the utterances triggered by placing rods as a Wittgensteinian language-game. Finally, I reflect on a 7½ minute video-clip of a 6-year-old interacting with the app.

Background: how to capture data on Froebelian block play for analysis?

The initial aim was to find a technical solution to a specific practical research problem: how to record children's arrangements, in play, of Froebel's Third Gift of eight, inch-wide wooden cubes (Froebel, 1895)? While there is a long history of studies of children's block play, technically many of them appeared to rely on adult observation and categorisation of block constructions, 'live' or filmed (Kersh et al., 2008), or more recently on specially-built cubes containing electronic devices to transmit their own positions and orientations (Watanabe et al., 2004). There was also evidence from computer vision research that software could recognise toy bricks in a video stream in near real-time, for example to sort and count Lego bricks (West, 2019). I was keen to explore whether software might be able to detect, from a video image, the positions and orientations of wooden cubes placed on a table, as coordinates, to facilitate both anonymised data gathering and statistical analysis. After some research, not finding any suitable apps commercially available, I contacted a computer vision programmer, PySource, and we collaborated on developing an app.

After several unsuccessful attempts to recognise cube positions from an overhead camera using traditional edge-detection algorithms, and colour detection by painting each cube a different colour, we decided to try a new approach, with two alterations: (1) to simplify the problem by switching from cubes to Cuisenaire rods, which are distinguishable by both their different lengths and their colours, and are typically laid flat rather than stacked; and (2) to experiment with artificial neural network algorithms to recognise the rods. The initial proof of concept successfully demonstrated the viability of the new approach in an app running on a standard laptop with a Graphics Processing Unit (GPU) chip and webcam, and we were fortunate to receive a seedcorn funding grant from the Bristol Digital Futures Institute which enabled us to develop the app and run a trial in collaboration with a primary school.

Like Froebel's cubes, Cuisenaire rods have a history of being used in diverse contexts in which their meaning is transformed, from numbers and lengths in mathematics education, to language learning, in which they may stand for objects, parts of speech, or phonetic sounds (Paipa, 2010). Thus rods appeared to offer similar affordances for research into diverse worldings comparable with Froebel's cubes. For a theoretical framework to understand how play with rods or blocks can generate diverse meanings, given the app could respond with a small set of words to a limited set of rod-placing actions, I turned to Wittgenstein's concept of language-games.

Wittgensteinian language-games as a tool to understand 'worldings' with blocks

For Froebel, when creating a world in play, a cube can become almost anything:

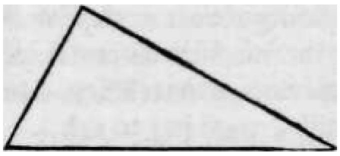
> So, for example, the cube can be now a table on which something is placed for the child. Again, it can be a stool on which the mother places her feet; again, a chair on which she sits with the child; again, the hearth on which something is to be cooked for the child. (Froebel, 1895, p.99)

It is notable in Froebel's transformations of the cube that he does not simply say the cube can be a table, a stool, a chair etc., but always adds a purposeful action, of someone placing something, sitting, cooking etc. – usually involving the child - as if an embodied action, a 'doing', or active role in a story, helps to evoke the worlds in which the cube's meaning is transformed.

For Wittgenstein, actions were likewise crucial in giving language context and meaning: "I shall also call the whole, consisting of language and the actions into which it is woven, the 'language-game'" (Wittgenstein, 1958, p.5). He gives as a prototypical illustration a primitive language-game involving actions on stone blocks:

> Let us imagine a … language is meant to serve for communication between a builder A and an assistant B. A is building with building stones: there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words "block", "pillar", "slab", "beam". A calls them out;—B brings the stone which he has learnt to bring at such-and-such a call (p.3)

One can imagine for a period the assistant might be sent back repeatedly until he learned to play the game with the builder of which block was called what. Perhaps the builder would call 'Slab!', the assistant would bring a 'wrong' stone, the builder would say 'Pillar?', shaking her head and pointing him away, before calling after him 'Slab!' When the correct stone is brought she might repeat 'Slab!' to endorse their entrained 'worlding'. Here there is no one 'true' meaning of each block, outside of the actions and language woven around it, the language-games people play with it. The same may be said of manipulatives and other objects in mathematics classrooms:



> The primitive language-game which children are taught needs no justification; attempts at justification need to be rejected. Take as an example the aspects of a triangle.
>
> This triangle can be seen as a triangular hole, as a solid, as a geometrical drawing; as standing on its base, as hanging from its apex; as a mountain, as a wedge, as an arrow or pointer, as an overturned object which is meant to stand on the shorter side of the right angle, as a half parallelogram, and as various other things. (Wittgenstein, 1958, p.5)

Some of these interpretations of a triangle come from language-games played in mathematics classrooms, others from elsewhere, however according to Wittgenstein when we start to interpret and contextualise the triangle in a certain way – when we start to play a particular language-game with it – in that moment that is what we see and what the triangle 'is', there is no further justification:

> …A triangle can really be standing up in one picture, be hanging in another, and can in a third be something that has fallen over.—That is, I who am looking at it say, not "It may also be something that has fallen over", but "That glass has fallen over and is lying there in fragments". This is how we react to the picture. (Wittgenstein, 1958, p.5)

How and why an image evokes different language-games is for Wittgenstein mysterious. In some examples, such as in textbooks, it may be that "text supplies the interpretation of the illustration," (1958, p.193) - suggesting the possibility that apps may similarly be able to supply verbal interpretations. In others, such as artworks, "I think custom and upbringing have a hand in this" (p.201).

In mathematics education, too, it is sometimes mysterious how and why some children learn to join in the language-games played in classrooms easily, while others do not, and contextual language, custom and upbringing may also be important. For example, in many primary classrooms in England following an 'mastery' approach there is now a focus on developing skills in bar modelling, following a Concrete-Pictorial-Abstract sequence of stages, in which each stage has its own language-game of words and actions on manipulatives, which may be supported by textbooks as well as the teacher's own language and actions (NCETM, 2021).

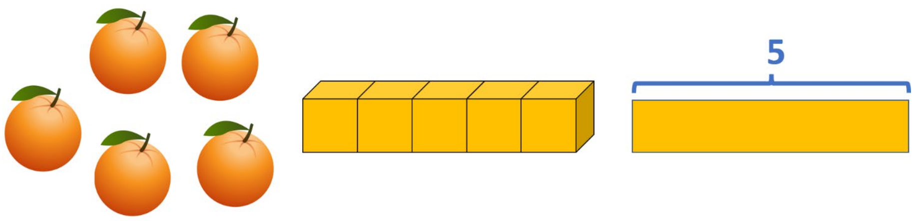


Figure 1. Typical stages in bar modelling: from counting a group, to aligning blocks, to labelling a bar

Typically, first 'real' objects such as fruit (or models or pictures of fruit) are grouped and counted as part of a story or 'word problem'. Then similar language is used around the action of grouping and counting interlocking cubes. These are then lined up and connected into a solid row, which can now be said to be a number 'long' and lengths can be compared or added. Finally, a diagrammatic 'bar model' of the conjoined blocks can be drawn on paper (Figure 1). In some classrooms Cuisenaire rods may be used in parallel as tangible, three-dimensional bar models (ATM, 2021).

This step-by-step transformation of the language-game from handling five oranges, for example, to drawing a rectangle labelled '5' is carefully managed by the teacher, usually over a period of months or years at the start of primary school. Because of the distinct actions and language used at each stage, they can each be conceived as separate language-games with simple rules, similar to Wittgenstein's building site. And it is these rules – interweavings of language and actions – the children are learning in order to play the new language-games at each stage.

Drawing on this instance of children starting to act on and talk of bars as lengths in bar modelling, the experiment design can thus be framed as way of endorsing a new language-game, in a similar way to the builder teaching the assistant. The number of oranges may stay the same, but now instead of the action of pointing to each in turn to count them, a bar may be drawn diagrammatically and labelled with a length of '5' to invoke the new language-game. Or analogously, a Cuisenaire rod may be placed and the word 'five' spoken.

Experiment design

In collaboration with the programmer a prototype app was developed in Python, using the Mask-RCNN artificial neural network algorithm to recognise Cuisenaire rods from a webcam feed and generate pre-programmed sound files accordingly, for example the lengths of each rod placed. An apparatus incorporating a webcam connected to a speaker-equipped laptop running the app was designed to be set up easily on a school desk for safe, supervised use by young primary school children.

A group of twelve 5- and 6-year-olds in year 1 of a local primary school in south-west England were recruited, ensuring socio-economic, ethnic and gender diversity and inclusion of children with special educational needs and disabilities. Over the period of two, 90-minute sessions in July 2021, closely supervised by a researcher, the children were invited singly or in pairs, for 10-15 minutes at a time, to try out a 'game' which involved making the computer say words suggested by the researcher, such as the number three, or the two times table, by placing rods on the desk, and taking their hands away to make it speak.

The app captured snapshots of the rods, and a video camera captured the audio generated. All video was edited to anonymise it. As the first time the prototype app had been trialled in the field, this was also an early technical trial of the technology. In this paper I focus on a 7½-minute video clip of one 6-year-old interacting with the app, and a set of ten Cuisenaire rods – one of each colour and length.

Results

In practice there were many technical challenges to running the prototype object-detection app in the field for the first time, for example varying lighting conditions when the sun came out affecting recognition, movement of the table as children leaned on it, and the app struggling to recognise unexpectedly large piles of rods, which were all valuable learnings to feed into technical development of future iterations. The researcher was not only providing prompts for the game as planned but also troubleshooting technical issues while attempting to keep the 'game' going.

Nevertheless, there were several sessions which ran relatively smoothly technically, such as the one analysed below, albeit with some occasional management of rod detection issues. For conciseness the dialogue in Figure 2 is summarised, but below is a typical excerpt to give a sense of the full transcript:

> Researcher: OK I think it's actually when you're wobbling the table slightly it sort of starts again when it sees a movement.
>
> So how about – you're doing really well - how about... do you think you can make it say the number... What number do you think you can make it say?

Participant: Ten?

Researcher: OK, do you want to try to make it say ten?

Figure 2 represents all the distinct rod arrangements made by the participant in response to the prompts during the 7½ minute clip, with the numbers then spoken by the app below. The numbers in brackets e.g. '(6)' indicate when that rod/number was occasionally not detected and so not spoken by the app.

In reviewing the correlation between the prompts and the number words generated by the participant's block placements via the app, the participant's understanding of the rules of the language-game – what actions triggered which of the app's utterances - appeared to change over the course of the 7½ minute session, such that it increasingly matched. Three phases were identified:

- An initial **phase A** when the actions generated app utterances which only occasionally matched the prompt. Multiple rods were arranged to generate a single expected utterance.
- A **phase B** when there was some consistent placing of a rod that generated the expected utterance, alongside other rods. And some returning to reiterate previous placements.
- A third final **phase C** in which rods were placed which mostly generated the expected utterances. Where there was a discrepancy noticed, a single rod – the last placed – was replaced with one longer or shorter according to the perceived error until the expected utterance was made.

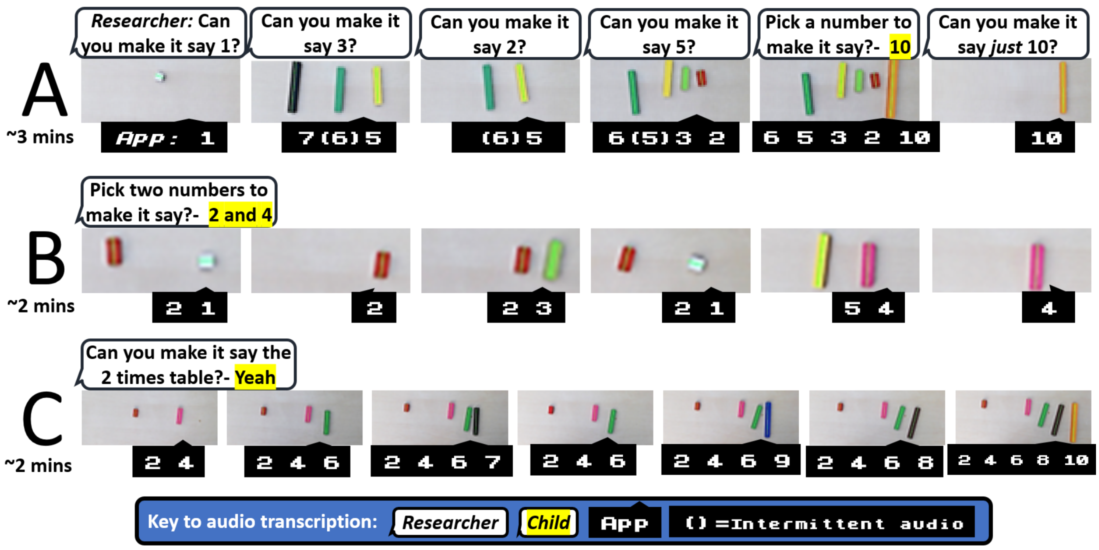


Figure 2: snapshots from a ~7-minute video clip of a 6-year old's rod placements and interactions with the app, illustrating a shift from rods as counters (A), to experimentation (B), to rods as lengths (C).

Discussion: From fingers to rods

In terms of language-games, the first phase visually resembles holding up fingers corresponding to an utterance, for example three fingers to respond to the word 'three'. These rod 'fingers' can be seen as counters, with little attention paid to which fingers, or their lengths. The final phase resembles the language-game of relating utterances to the lengths of the rods. The second, intermediate phase appears to be unstable, with elements of both the first and last phases, as if seeing that familiar rules do not apply, new rules for the new language-game must be conjectured and tested.

If one understands education, and mathematics education in particular, as developing expertise in certain context-specific language-games, then the results indicate that a 6-year-old may be able to adapt their actions and interpretations to participate in a new language-game, over the period of a few minutes. In the theoretical framework described above, this can be understood as a change in interpretation of the rods, a 're-seeing', or 're-worlding' of the rods, such that they stop resembling fingers or counters, and start becoming Cuisenaire rods with meaningful lengths and colours. A similar shift from the language and actions of counting to the language and actions of lengths can be seen in stages of bar modelling.

Clearly as an early prototype much work is required to make the app apparatus and software more robust and portable to be usable in any larger-scale research. And to analyse play with Froebelian cubes the app would need to be developed to recognise cubes stacked in three dimensions. However, as available algorithms and consumer-level chip processing power improve, it seems increasingly feasible to develop these and other functionalities as tools for future research.

Acknowledgement

This research was funded by a grant from the Bristol Digital Futures Institute. I thank collaborators Professor Alf Coles, software developer PySource and the children and staff of the participating primary school.

References

ATM (2021) *Cuisenaire – from Early Years to Adult*. https://www.atm.org.uk/Shop/Cuisenaire---from-Early-Years-to-Adult-Book-and-download/ACT103pk

Froebel, F. (1895). *Friedrich Froebel's pedagogics of the kindergarten* (Vol. 30). D. Appleton.

Kersh, J., Casey, B., & Mercer Young, J. (2008). Research on spatial skills and block building in girls and boys: The relationship to later mathematics learning. In B. Spodak and O. Saracho (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 233–253). Information Age Publishing.

NCETM (2021). *The bar model.* https://www.ncetm.org.uk/classroom-resources/ca-the-bar-model/

Paipa, K. (2010). Te whakapapa o te reo i roto i te whanau. *Māori And Indigenous Review, 3*, 1–15.

Watanabe, R., Itoh, Y., Asai, M., Kitamura, Y., Kishino, F., & Kikuchi, H. (2004). The soul of ActiveCube: implementing a flexible, multimodal, three-dimensional spatial tangible interface. In *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology.* Association for Computing Machinery.

West (2019). *How I created over 100,000 labeled LEGO training images*. https://towardsdatascience.com/how-i-created-over-100-000-labeled-lego-training-images-ec74191bb4ef

Wittgenstein, L. (1958). *Philosophical investigations*. Macmillan.