Radboud University

The Predictive Processing Account of Infant Colour Vision Development on Generative Models

Bachelor's Thesis Artificial Intelligence

Author: B. M. W. Waelbers [a]
Student Number : 4591704
Email Address : b.waelbers@student.ru.nl

Department of Artificial Intelligence
Radboud University Nijmegen

Under supervision of:
 Dr. J.H.P. Kwisthout [a,b]
 D. Rutar M.A. [a,b]

[a] Artificial Intelligence
[b] Donders Institute for Brain, Cognition and Behaviour

Abstract

Predictive processing is a theory that models how the human brain tries to predict sensory input from the environment. How generative models are being shaped in infancy is still a new, but very interesting topic. In my thesis, I focus on *developmental* predictive processing by investigating the conceptual and behavioural consequences of infant development on generative models. In this project, I specifically look at the development of colour vision in new-borns. I modelled two different scenarios, one in which intensities are learned based on movement and where colour perception is added afterwards, and one in which the model immediately learns intensity and colour based on movement. My research question is based on both scenarios: "What is the difference in the size of the prediction error between step-wise learning intensity and colour perception based on movement (Scenario 1), and immediately learning intensity and colour perception based on motor movement (Scenario 2)?". Based on the fact that children start by only perceiving intensities, and later learn to discriminate between colours, my expectation was that the first scenario results in a lower total size of prediction error. I compared both scenarios based on the total size of the prediction error computed by the Kullback-Leibler divergence. K-means clustering resulted in two divisions. In the first division, both prediction errors were similar, so no scenario performed better. For the second division, learning intensity and colour directly from motor movement resulted in a lower prediction error.

Keywords: predictive processing, generative models, child development, colour vision

Introduction

Predictive processing is nowadays seen as a leading theory on how the human brain processes sensory inputs from the environment, by predicting the inputs based on generative models and processing only that part of the input that could not be predicted (Clark, 2013). Predictive processing offers a possible explanation of how the brain is organized (Clark, 2013). For example, Seth (2014) describes a predictive processing theory of sensorimotor contingencies. The theory provides explanation for the occurrence of perceptual presence in normal perception, just as its absence in synaesthesia (Seth, 2014). Similarly, Perfors and colleagues (2011) use Hierarchical Bayesian models that offer an alternative computational level explanation of the development of learning object names and categories. Predictive processing is a good explanation of how low level perception is implemented in the infant brain. Now, the challenge is to provide a theory that it is also able to explain how higher cognitive functions are developed in infancy.
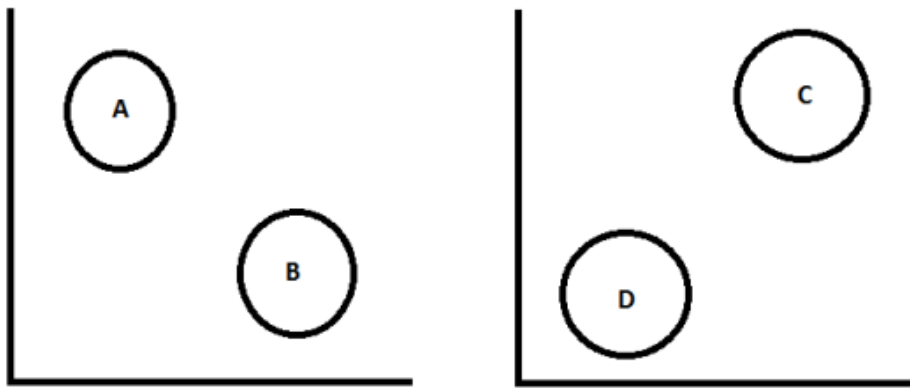
Predictive processing

In predictive processing, a hierarchical generative model is used to minimize the prediction error (Clark, 2013). Every layer of the hierarchical model can be represented as a causal Bayesian network

(Kwisthout, Bekkering, & Rooij, 2017). Those networks are probabilistic graphical models that represent stochastic dependencies between variables. Every layer has three kinds of variables: hypothesis variables (Hyp), prediction variables (Pred) and intermediate variables (Int) (Kwisthout, Bekkering, & Rooij, 2017). The arcs between the variables depict causal relations (Kwisthout, Bekkering, & Rooij, 2017).

Generative models are initially low in detail. This means that predictions based on these models are very global, resulting in low prediction errors and almost no informational content. The next step is to fine-grain the model, which will induce more prediction errors, because the prediction will be more detailed. Training the models results in the fact that the model is refined, so the predictions will be more fine-grained and the prediction errors will decrease (Kwisthout et al., 2016). Kwisthout and colleagues (2017) discuss the example of a die. When the model is low in detail, the number of a die is classified as 'odd' or 'even'. The probability of being in one of the categories is 1/2, which is quite high, but there is little information about the outcome of a rolled die. When the model is more precise, the result of a die can be classified by number of dots on the die. The probability for a fair die is 1/6 for side of a die. There is more information, but this will result in more prediction errors. This trade-off between precision and information gain is an important topic in predictive processing, but will not be discussed in this research.

The model tries to predict sensory inputs, to compare these predictions to the actual observation. This comparison is performed at each level of the hierarchy (Kwisthout, Bekkering, & Rooij, 2017). The error is the difference between the prediction and the observation. The size of the prediction error is computed by the Kullback-Leibler divergence between the two distributions. The goal is to make the prediction as close to the actual observation as possible. If the models were perfect, the predicted input would be the same as the actual observation. To get the prediction as close as possible to the actual outcome, the prediction error should be lowered. Prior knowledge is combined with new observations to update the existing model. There are two ways to lower the prediction error: by improving the generative model, or by intervention in the world (Kwisthout, Bekkering, & Rooij, 2017).

In her BSc thesis, Ter Borg (2017) proposed a generative model of predictive processing by individual experience. She used a LEGO Mindstorms NXT robot. The robot was put at a certain spot in a dark environment with a light source, facing the dark side of the environment. The wheels of the robot rotated based on motor commands, resulting in a change of light intensity. She clustered the motor and sensory input of the robot to make a probability distribution based on these experiences, see Figure 1. The generative model predicted the light intensity based on the motor commands, using k-means clustering. She used a binary causal Bayesian network with two variables, namely a prior (P[a] and P[b]) and a conditional probability distribution.

Figure 1 Generative model of predictive processing by individual experience; clustering based on motor movement and intensity measurement (Ter Borg, 2017). The left figure shows the expected clustering of the motor movement, plotting wheel 2 to wheel 1. The right figure shows the expected intensity clusters, plotting intensity 2 on intensity 1. Cluster A is the motor cluster for turning left, cluster B for turning right. Cluster C labels the high intensity trials and cluster D contains the low intensity trials.

The main idea of her thesis is very interesting, so I decided I wanted to expand on her thesis. I used her work as the starting point of my own. There are a couple of options to enlarge her research, based on the possibilities of lowering prediction error (as discussed earlier) and contributing to the predictive processing theory. The first one is relearning the clusters, by adding a third cluster and to see how the model reacts to this change. Another expansion is to learn contextual modulation, for example by adding a wall that blocks the light, which results in different light intensity inputs. A third way of building on Ter Borg's research would be to explore the developmental aspect of her research and thus provide a developmental account of the predictive processing theory. One way of studying developmental consequences is by examining physical changes in the robot. Examples of these physical changes in the robot are adding sensors or motors or making the robot heavier or taller. I chose the third option of contributing to the theory, by adding a colour sensor to the robot. This addition is equivalent to the development of colour vision in infants, which I will discuss in the next section.

Colour vision development in infants

In the ongoing literature, it is commonly accepted that new-borns are only capable of differentiating between the brightness of an object. However, there is some disagreement about when they start to develop colour visions. There are some studies that show that infants do have colour vision, which means they can differentiate between chromas[1]. Wilton (1937) showed that infants between 15 and 70 days do have the ability to discriminate between colour combinations. More specifically, they found a significant difference between colour and brightness for the combinations of red, yellow-green, green,

---

[1] Chroma : the quality of a colour's purity, intensity or saturation

and blue-green. Peeles and Teller (1975) showed that two-month-old human infants do have some form of colour vision. Children were able to discriminate between a red and a white bar, suggesting that infants of two months of age are at least dichromatic. A study showing the opposite results is the research of Clavadetscher and colleagues (1988). They tested 3- and 7-week-old infants for their chromatic discrimination. The 3-week-old children were not able to make chromatic discriminations. The 7-month-old children did show a difference between chromas, but this was still far from perfect. Adams and colleagues (1987) showed that infants with the age under three months show a different response to different chromas than adults. Adults have a different response, since they can discriminate between colours and chromas, while three-month-old infants cannot.

For this thesis, the exact age at which children have fully developed colour perception does not really matter. It is enough to know that new-borns only perceive the differences in brightness of a colour and that they develop colour perception while growing older (Adams, 1987). Because there is no research that showed colour vision in infants under the age of 15 days, I will use this threshold when talking about mimicking colour vision in robots.

Developmental robotics

Developmental robotics is a recent domain that lays the link between behavioural research and robotic modelling. It combines the theory, robot programming and behavioural explanations in one research program. This combination is made to identify the ambiguities in the real world, exactly the ones for which computer simulations would make an assumption about the environment (Deák, Fasel, & Movellan, 2001). Robotic modelling allows the comparison of computational theories. Working with robots allows one to study internal representations and processes. This is in contrast with behavioural researches, who only look at the behaviour of people while trying to study internal states (Deák, Fasel, & Movellan, 2001). The robotic aspect helps to explain behaviour, instead of only observing it.
Cognitive developmental robotics is one step further than developmental robotics. It focuses on higher cognitive functions. Other agents and the interaction between agents are the central point of this domain. Body representation and social behaviour are hypothesized, meaning that artificial systems try to mimic higher order processes and interaction between robots (Asada et al., 2009). This is what will be possible in the future and adds to the background of developmental robotics, but is too complicated for this thesis.

A new theory in this field is the robo-havioural methodology by Otworowska and colleagues (2015), which is the intermediate step between theoretical and empirical investigation. The FOES methodology is an example of a concrete way to lay a link between behavioural research and robotics modelling. FOES is the abbreviation of Formalize, Operationalize, Explore and Study (Otworowska et

al., 2015). F stands for formalizing verbal theories into computational models which is implemented (O) for a working robot. The consequences of various design choices and parameter settings are explored (E) to generate empirically testable hypotheses. These hypotheses are studied (S) in behavioural or imaging experiments. The goal of this method is not to build smart robots, but to see where the gaps and ambiguities of the theories are and what the consequences of assumptions and/or design choices are (Otworowska et al., 2015).

In this thesis, I will combine the developmental robotics account together with the robo-havioural methodology and use the robot to test the theory of colour vision development in infants. The goal of this thesis is to expand the existing theory and to research developmental predictive processing by investigating the consequences of colour vision development in new-borns on generative models. There is a lot of literature about predictive processing, just like infant development is a widely discussed topic. When combining these two domains, there is still a gap, which I will partly try to close, using the robo-havioural methodology and a LEGO Mindstorms robot.

15 days old new-borns only react to brightness and light intensity of objects, which means that they only see the difference in light intensity. This can be modelled in a robot by solely using a light intensity sensor, building up the model by using just the limited sensory information. Children will start learning about the world, just like the robot will learn about its environment. While the children grow, they start to develop perception of colours. The child has to adjust the perceptual model that she has learned in the past days, to include the colour sensation. The development of vision from light/dark to colour is mimicked by adding a colour sensor to the existing model. Now, there are two different ways of adding this colour sensor. The first approach aims to imitate how new-borns learn the world: by originally starting with the model of the light sensor and afterwards adding the data obtained by the colour sensor to refine the model (Scenario 1, Figure 2). The second way is to train the model by immediately using the light sensor data and the colour sensor data, based on the motor movement (Scenario 2, Figure 3). The algorithm of this scenario is smaller and is easier to implement. Still, the state space to be learned will be larger and can have an influence on the size of the prediction error. Therefore, my research question is: "What is the difference in the size of the prediction error between step-wise learning intensity and colour perception based on movement (Scenario 1), and immediately learning intensity and colour perception based on motor movement (Scenario 2)?". Based on the developmental research, I expect the first scenario to have lower total size of prediction error.
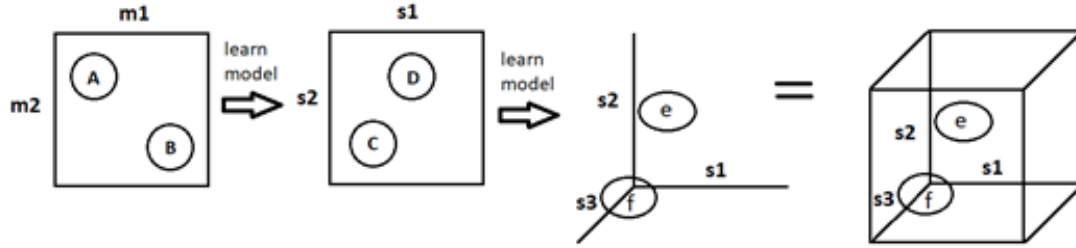
Scenario1:

Figure 1 Scenario 1: developing the model by first learning the intensity clusters based on the motor movement clusters, and afterwards adding the data of the colour sensor to cluster the data. The first coordinate system plots the rotation of wheel one to the rotation of wheel two in a 2D-space. The second coordinate system plots the intensities of the sensors in a 2D-space. The third coordinate system shows the same clusters as coordinate system 2, but it shows the 3D data. The last coordinate system shows the clusters based on colour measurement and intensities in a 3D space. M1 and m2 are the inputs of the model, namely the rotation of the wheels. S1 is the input from the first light intensity sensor. S2 is the input from the second light intensity sensory. S3 is the input from the colour sensor.
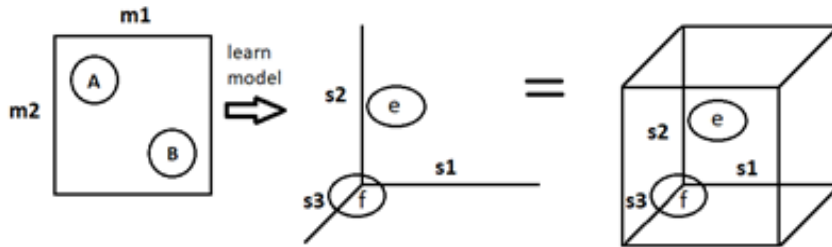
Scenario2:

Figure 2 Scenario 2: developing the model by immediately clustering the intensity and colour data based on the motor movement clusters. The first coordinate system plots the rotation of wheel one to the rotation of wheel two in a 2D-space. The second (which is the same as the third) coordinate system shows the data clustered based on the intensities and the colour measurement in a 3D space. M1 and m2 are the inputs of the model, namely the rotation of the wheels. S1 is the input from the first light intensity sensor. S2 is the input from the second light intensity sensory. S3 is the input from the colour sensor.

Methods

Robot

For this research, I use a Lego Mindstorms EV3 robot. My robot needs to move and has to measure colour and intensity. Therefore, my robot has two motors and two sensors, see Figure 4. Each motor is connected with a wheel. Motor A is connected with the right wheel and Motor D is connected with the left wheel. To mimic the vision of infants, the first sensor (port S1) is placed on top of the robot. In this way, it imitates the eyes an infant best. This sensor measures ambient light, which corresponds to the intensity of the environment. The second sensor (port S2) is able to measure both intensity of the environment and the colour of the floor. The colour sensor is not very sensitive: it only detects the colour of an object when it is within two centimetres. Because of this, the sensor is placed in front of the robot, facing down, about one centimetre of the ground. In this way, it is able to measure colours painted on the floor.
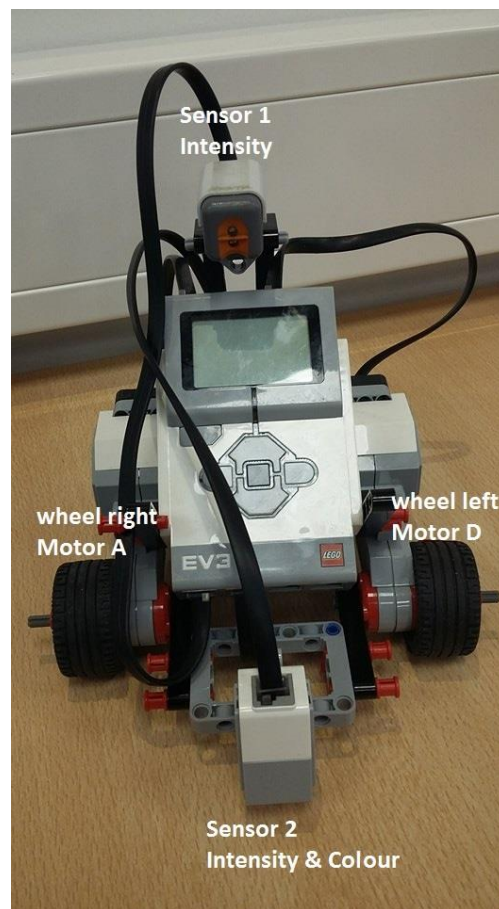
Figure 3 Robot Design: the Lego Mindstorms EV3 robot has two motors, each connected to a corresponding wheel, and two sensors. The first sensor only measures intensity, while the second sensor measures both intensity and colour. The left wheel is connected to motor port D, and the right wheel is connected to motor port A.

Environment

The environment of the robot consists of a light source and a plate to rotate on, see Figure 5. The light source is placed a bit higher than the robot, see Figure 6. When the light was placed on the same level as the robot, the sensors gave a very low intensity value until the robot was almost facing the light. This resulted in the same value for almost half a round, so I placed the light source higher, resulting in a more variable light intensity. The size of the plate is 30cm×30cm to create enough space for the robot to completely turn around without falling off.
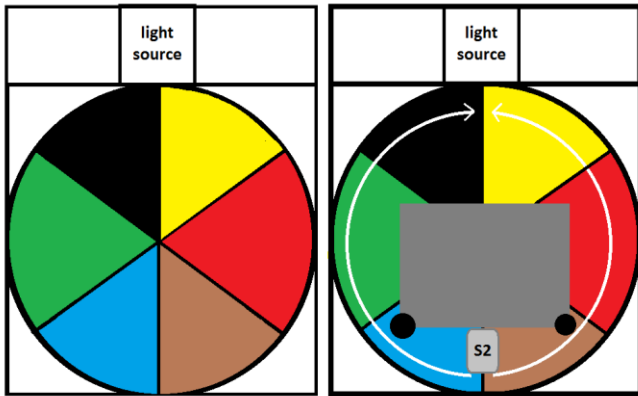
Figure 4 Schematic picture of the robot's environment. The pie-shaped plate is divided in 6 parts, showing the low intensity colours (black, green, blue) left and the high intensity colours (yellow, red, brown) right.
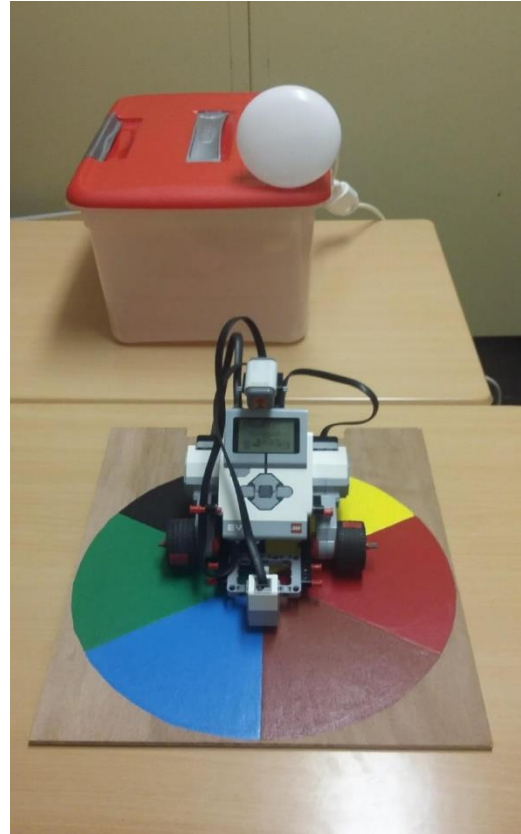

Figure 5 Picture of the environment of the robot, with the light source and the plate the robot turns on.

An EV3 robot is able to discriminate between colour and no colour. It can distinguish between seven different colours: black, blue, green, yellow, red, white and brown. Therefore, the plate on which the robot is turning is divided in six different colours, in the shape of a pie chart. The colours are six of the seven colours the EV3 sensor can recognize, namely black, green, blue, brown, red, and yellow. Every colour has the same surface, with an angle of 60°. The colours are divided based on intensity: the left ones (blue, green, black) have a low intensity, the right ones (brown, yellow, red) have a high intensity, see Figure 5. Because of this division, every colour corresponds to a different movement and a different light intensity. The intensities measured by sensor 2 do not depend much on the light intensity: they depend mostly on the intensity of the colour. The mean intensity of a colour depends on the brightness of the colour, see Table 1. Every colour has a colour ID, which is used in the *Results* section. Since white (colour ID number 6) is not in the pie chart, only ID 1,2,3,4,5 and 7 are used, see Table 1.

	Colour	Intensity Sensor2	Colour ID
turn right	black	0.03	1
	green	0.03	3

	blue	0.05	2
turn left	brown	0.18	7
	red	0.23	5
	yellow	0.46	4

Table 1 Colour Intensity & ID

Pilot study

A pilot study was performed to see what the environment should look like to make the best possible use of the colour sensor. There were three options: a coloured light, coloured tape, and paint. The coloured light did not work, since the sensor measures the actual colour of an object and not the colour of the reflected light. For the tape and paint, there was no exceeding method based on colour recognition. Their colours were both measured quite flawlessly, yet blue was a struggle for both. Eventually, the paint gives the best accuracy. One disadvantage of the tape is that it can come off. Also, it is relatively hard to apply the tape smoothly on the ground without bubbles of air underneath it. The paint is easy to apply and assures that it will stay like it should. Because of this and the higher accuracy of the blue colour, the environment of the robot consists of a painted, wooden board on which the robot can turn.

One possible struggle was that the colours would not be recognized when the robot was facing away from the light. Fortunately, the sensor did not have a problem with recognizing the colours in a dark environment, since it has a light in the sensor itself. The light in the sensor makes the robot able to recognize the colours.

Experiment

The goal of the robot is driving towards the light. Therefore, it starts each time at the same point, facing away from the light source. The robot turns left or right for a random amount of degrees. It never turns more than half a circle, because if it would rotate more, it would turn away from the light source again.

Calculations of rotations

The distance between the wheels is 151mm (including the wheels) (see Figure 7), which means that the circumference is 948.76mm, when the robot would turn only one wheel.

c1=$2 \times \pi \times 151 \approx 948.76$
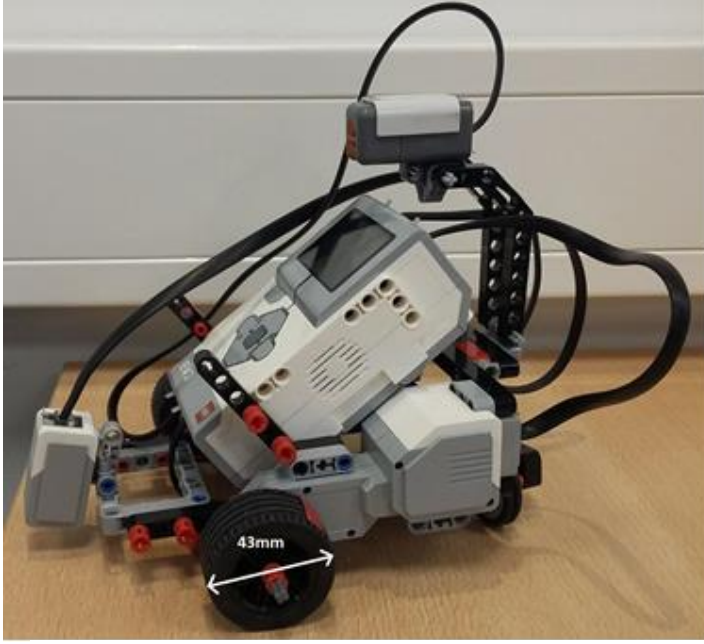
Figure 6 Wheel Diameter of the robot.
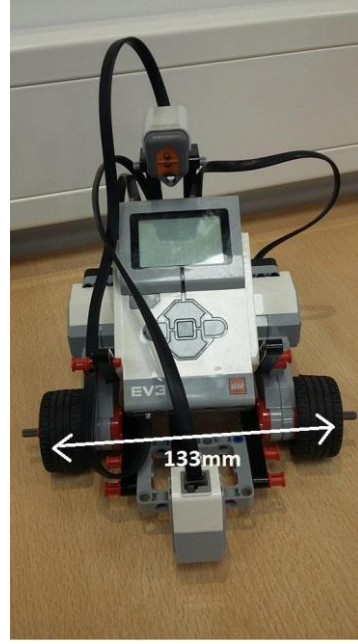
Figure 7 Diameter of the robot itself.

The diameter of the wheel of the robot (see Figure 8) is 43mm, resulting in a circumference of 135.1mm.

c2= $\pi \times 43 \approx 135.1$

The number of degrees one wheel has to turn to rotate a full circle is 2528°

d = $c_1/c_2 \times 360 = \frac{2 \times \pi \times 151}{\pi \times 43} 360 \approx 2528$

Because the robot only has to turn half a round and both wheels are going in the opposite direction, the maximum amount of degrees each wheel may rotate is 2528°/4 = 632° . This means the random value is a value between 0 and 632 .

After the rotation, the colour of the ground, intensity of sensor 1 and intensity of sensor 2 are measured respectively.

Analysing the data

I used k-means to cluster my data resulting from the robot experiment. The motor input, intensity output and colour output are all divided in different clusters.

First, I will explain what k-means is, why I used it in my research, and how I implemented it in MATLAB. Next, I will explain the conceptual model of my clusters.

Cluster analysis is an unsupervised learning algorithm, because there are no labels. Data without labels means that the data does not have defined categories or groups (Kumar, 2014). The algorithm is used for exploratory data analysis, to find hidden patterns or grouping in data. K-means is one of the most popular clustering algorithms (Kumar, 2014). The K-means algorithm performs as follows: Select k points as initial centroids. Then, form two clusters by assigning each point to its closest centroid. When you have k centroids, re-compute the centroid of each cluster. Repeat this until the centroids do not change. The algorithm used in this cluster uses $k = 2$, resulting in two clusters, since the motor movement is divided in two (turning left and turning right), and the intensity is based on low and high.

Because the goal is to find groups in the data while there are no labels available, clustering is the best method to use here. One of the reasons I used k-means is because it is a simple technique. Also, the number of clusters is fixed, so the algorithm should not choose the number of clusters itself. The value of k is in this research always *two*, because there is one binary cause variable and one binary effect variable.

MATLAB has a predefined k-means algorithm, which I used to perform the computations of clusters. The *kmeans(X, k)* algorithm in MATLAB performs k-means clustering to partition the observations of the n-by-p data matrix X into k clusters, and returns an n-by-1 vector containing cluster indices of each observation. It uses the *k-means++ algorithm* for cluster centre initialization, which uses a heuristic to find centroid seeds for k-means clustering. This centre initialization, proposed by Arthur and Vassilvitskii, is faster and gives a better output than the original k-means algorithm of Lloyd.

Clusters

Motor Clusters

Since the input of both scenarios is the same, the same motor clusters can be used. The wheels always rotate contrary, which means that if the left wheel is turning forward, the right wheel is turning backward and vice versa. To plot this, the rotation of the Motor A, so the right wheel is plotted on the horizontal axis and the rotation of the left wheel is plotted on the vertical axis, see Figure 9.
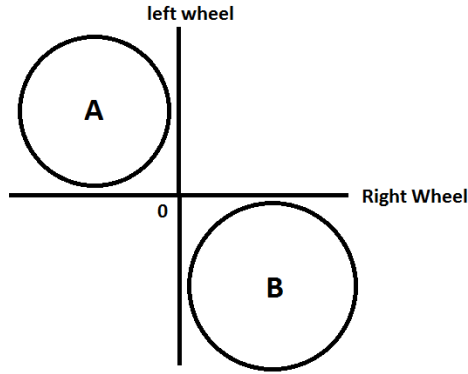
**Figure 8 Expected motor clusters. Both wheels rotate in contrary direction. The degree of rotation of the right wheel is plotted on the x-axis. The degree of rotation of the left wheel is plotted on the y-axis.**

Intensity Clusters

The intensity clusters are different from the motor clusters, due to the fact that intensity cannot have a negative value. There will be both high and low intensities, depending on the amount of rotation the robot is turning. So my expectation is that there is a cluster for high intensity values and a cluster for low intensity values, see Figure 10.
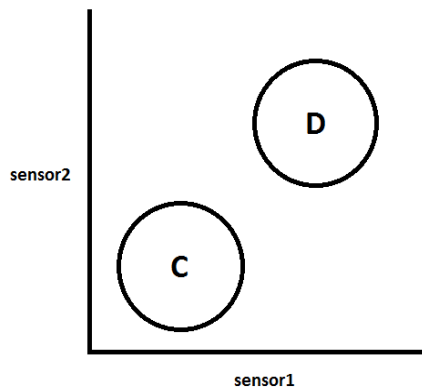
Figure 9 Expected intensity clusters. Sensor 1 is plotted on the x-axis, while sensor 2 is plotted on the y-axis. Cluster C represents trials with low intensity, cluster D contains trials with high intensity.

Colour Clusters

The colour clusters are two clusters in a three-dimensional space. This is because there are three variables plotted: the intensity of sensor 1, the intensity of sensor 2, and the colour ID. As explained in the *environment* part above, the colours are represented as numbers ranging from one till seven. These numbers are used to plot the clusters later on in this thesis. I expect to have two clusters that are clearly separated from each other, see Figure 11. This is based on the fact that the colours black, blue and green have low brightness, so the intensity of sensor 2 will be low too. The colours yellow, red, and brown have a high brightness, resulting in a high intensity for sensor 2.

14
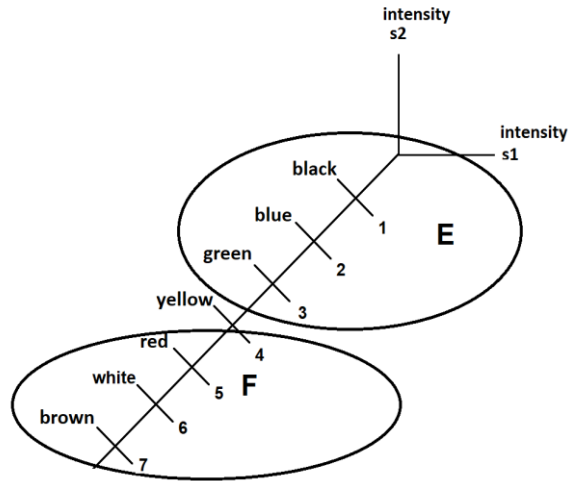
Figure 10 Expected colour clusters. The intensity measurement of sensor 1 is plotted on the x-axis, sensor 2 on the y-axis and the colour ID is plotted on the z-axis. Cluster E contains the samples with low intensity and low colour brightness. Cluster F clusters the trials with high intensity and high colour intensity.

For scenario 1, there are three steps: the first step is clustering the random motor inputs, the second step is clustering the intensity of sensor 1 and sensor 2, and the last one is clustering the intensities of sensor 1 and 2 and the colour sensor data, see Figure 12. All these clusters are used later to make up the model. For scenario 2, the second step of scenario 1 is left out. It only clusters the motor inputs and the intensities of sensors 1 and 2 together with the colour sensor data. Both scenarios are plotted in Figure 12.

Scenario 1:
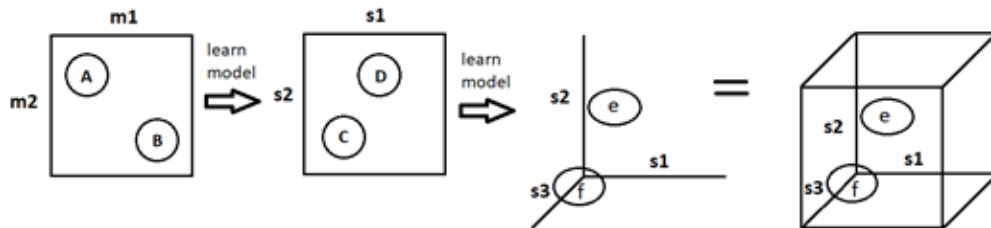
Scenario 2:
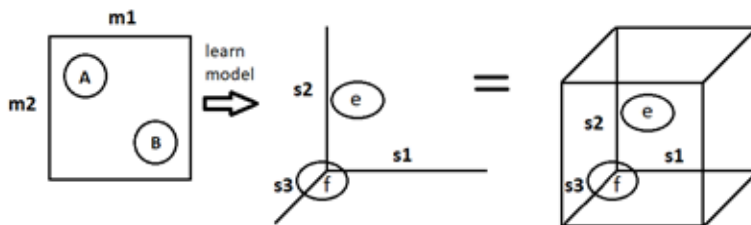
Figure 11 Model Scenario 1 & 2. Scenario 1: developing the model by first learning the intensity clusters based on the motor movement clusters, and afterwards adding the data of the colour sensor to cluster the data. Scenario 2: developing the model by immediately cluster the intensity data and colour data based on the motor movement clusters.

Results

In this section, I will give an overview of the data I collected in my robot experiment. Then, I will give the results of the clustering in some plots. Next, the predictive processing part is computed, starting by giving a short explanation of the Kullback-Leibler divergence. Afterwards, I will present the derived probability distributions and prediction errors based on that KL-divergence.

Data collection

Table 2 shows the final data. The full tables can be found in Appendix 1 and Appendix 2. For every trial, my robot gets two values: a value for the left wheel and a value for the right wheel. This value is the number of degrees that a wheel rotates. In the first 50 trials, the left wheel has a negative amount of degrees, resulting in a backward movement of that wheel, so the robot turns right. In the last 50 trials, the right wheel gets a negative amount of degrees, resulting in a turn to the left.
After the robot has finished turning, the robot's sensors measure both the intensity of sensor 1 and sensor 2, and sensor 2 measures the colour that it is facing. Colour ID is the number of the colour, which is discussed before in *environment*. Intensity 1 is the intensity measured by sensor 1. Intensity 2 is the intensity measured by sensor 2.

When all trials are completed, the data is clustered in MATLAB. By performing the k-means algorithm on the different subparts of the data, every trial is part of three clusters: a motor cluster, an intensity cluster and a colour cluster. The "1" stands for first cluster and "2" stands for the second cluster. The number of the cluster is irrelevant, it just shows that all trials with cluster number 1 are grouped together and that all trials with cluster number 2 are grouped together. The entire table with the data can be found in Appendix 1 and Appendix 2.

Trial	Degrees Left Wheel	Degrees Right Wheel	Colour ID	Intensity 1	Intensity 2	Motor Cluster	Intensity Cluster	Colour Cluster
1	-331	307	5	0,28986	0,26	1	1	2
2	-567	505	4	0,5379	0,45	1	2	2
…	…	…	…	…	…	…	…	…
50	-415	215	5	0,2899	0,28	1	1	2
51	119	-444	6	0,2733	0,09	2	1	2

…	…	…	…	…	…	…	…	…
100	167	-385	2	0,2791	0,05	2	1	1

Table 2 Final data. The rows contain the results per trial of the robot. The columns contain the degrees that the wheels rotate, the intensities and colours that are measured and the clusters (Motor, Intensity and Colour) each trial is in.

Data Cluster Plot

The clusters that result from the MATLAB k-means function can be plotted to give a clear overview of how the data looks like when it is clustered.

As discussed earlier, the motor cluster part can be used for both scenarios. The random values of both motors are plotted against each other, see Figure 13. The degrees of rotation of motor A are plotted on the x-axis, the degrees of rotation of motor D are plotted on the y-axis. Cluster A is given by the blue 'x' and cluster B is marked as a red dot.
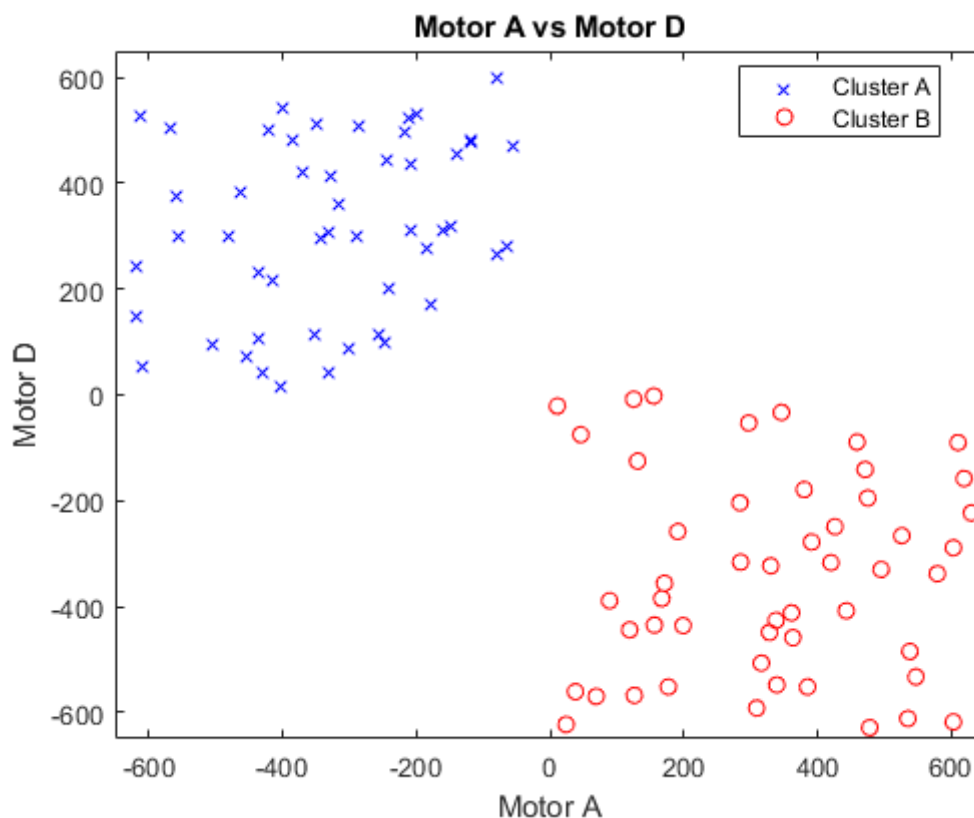
Figure 12 Plot Motor Clusters. Blue crosses are cluster A and red dots are cluster B. The rotations of motor A are on the x-axis, the rotations of motor D are on the y-axis.

For scenario 1, first the sensor values are plotted against each other, see Figure 14. The intensity values of sensor 1 are plotted on the x-axis. The intensity measurements of sensor 2 are plotted on the y-axis. Cluster C is given by the blue crosses, cluster D is given by the red dots.
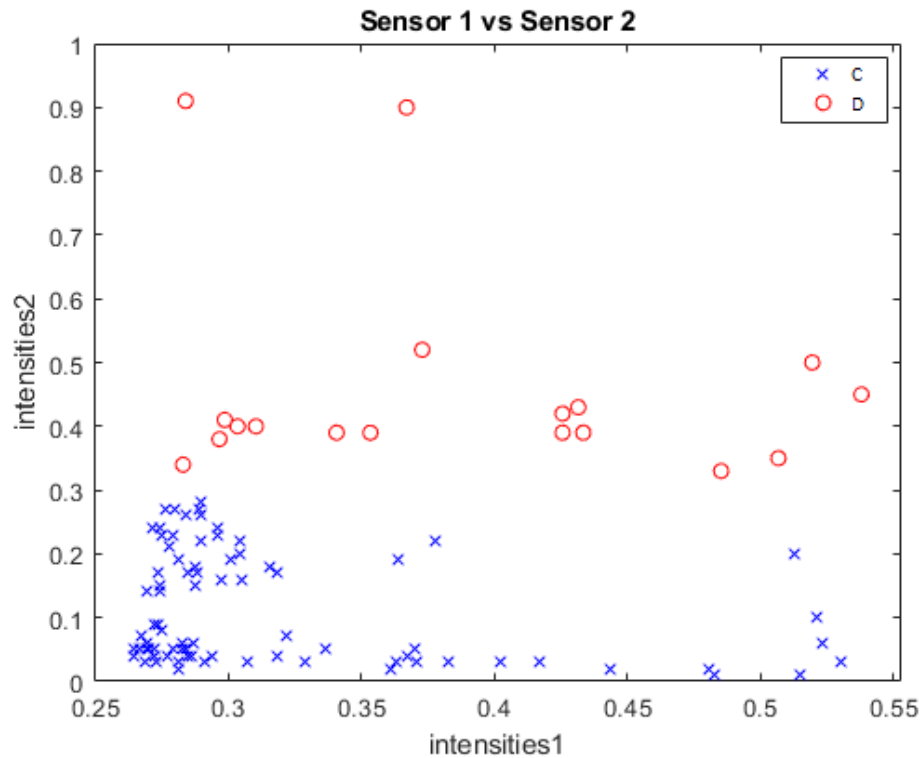
Figure 13 Plot Intensity Clusters. Blue is cluster C and red is cluster D. The intensities of sensor 1 are on the x-axis, the intensities of sensor 2 are on the y-axis.

The second step is to plot both intensity values against the colour ID. This was different per run, resulting in two different divisions. The first division is visible in Figure 15. The intensity of sensor 1 is plotted on the y-axis (right axis), the intensity of sensor 2 is shown on the x-axis (left axis) and the colour ID is plotted on the z-axis (vertical axis). Cluster E contains the samples with high intensity/high colour intensity with a blue 'x' and cluster F marks the trials with low intensity with a red dot.
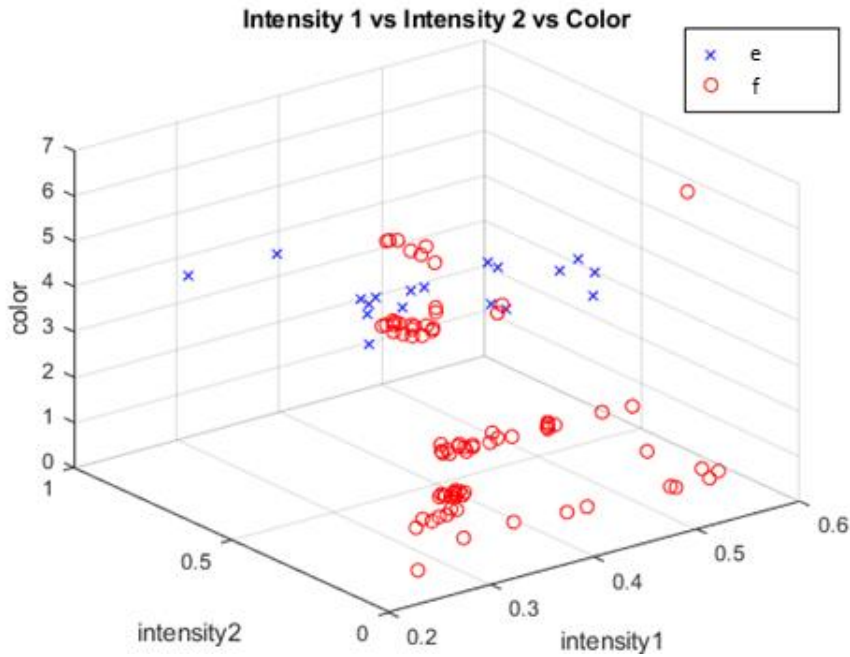
Figure 14 Plot Colour Clusters Scenario 1, Division 1. The intensity of sensor 1 is plotted on the y-axis (right axis), the intensity of sensor 2 is shown on the x-axis (left axis) and the colour ID is plotted on the z-axis (vertical axis). Cluster E contains the samples with high intensity/high colour intensity with a blue 'x' and cluster F marks the trials with low intensity with a red dot.

The second division is visible in Figure 16. The intensity of sensor 1 is plotted on the y-axis (right axis), the intensity of sensor 2 is shown on the x-axis (left axis) and the colour ID is plotted on the z-axis (vertical axis). Cluster E contains the samples with high (colour) intensity with a red dot and cluster F marks the trials with low intensity with a blue 'x'.

Figure 15 Plot Colour Cluster Scenario 1, Division 2. The intensity of sensor 1 is plotted on the y-axis (bottom right axis), the intensity of sensor 2 is shown on the x-axis (bottom left axis) and the colour ID is plotted on the z-axis (vertical axis). Cluster E contains the samples with high (colour) intensity with a red dot and cluster F marks the trials with low intensity with a blue 'x'.

In scenario 2, there were two different outcomes when running the algorithm multiple times.

The first clustering division is visible in Figure 17. The intensity of sensor 1 is plotted on the y-axis (right axis), the intensity of sensor 2 is shown on the x-axis (left axis) and the colour ID is plotted on the z-axis (vertical axis). Cluster E contains the samples with high (colour) intensity with a blue 'x' and cluster F marks the trials with low intensity with a red dot.

Figure 16 Plot Colour Cluster Scenario 2, Division 1. The intensity of sensor 1 is plotted on the y-axis (right axis), the intensity of sensor 2 is shown on the x-axis (left axis) and the colour ID is plotted on the z-axis (vertical axis). Cluster E contains the samples with high (colour) intensity with a blue 'x' and cluster F marks the trials with low intensity with a red dot.

The second clustering division can be seen in Figure 18. The intensity of sensor 1 is plotted on the y-axis (right axis), the intensity of sensor 2 is shown on the x-axis (left axis) and the colour ID is plotted on the z-axis (vertical axis). Cluster E contains the samples with high (colour) intensity with a blue 'x' and cluster F marks the trials with low intensity with a red dot.
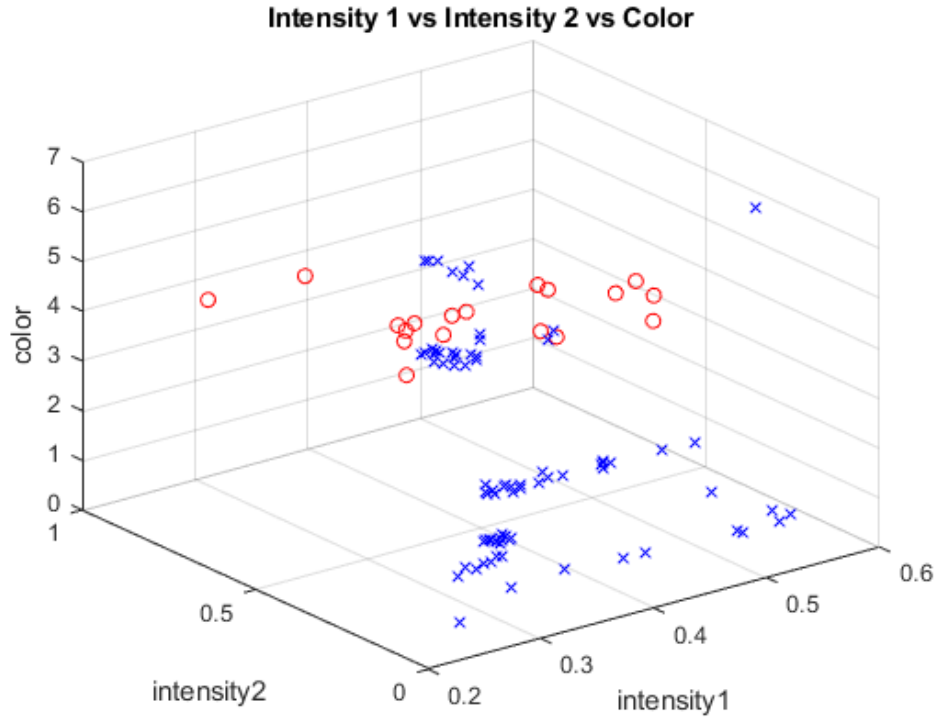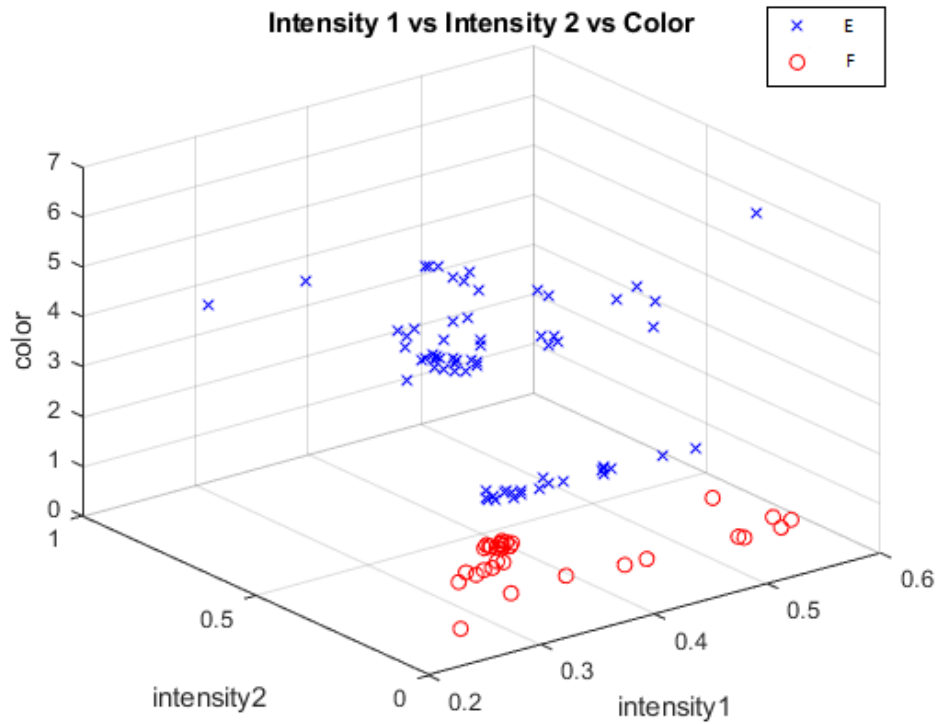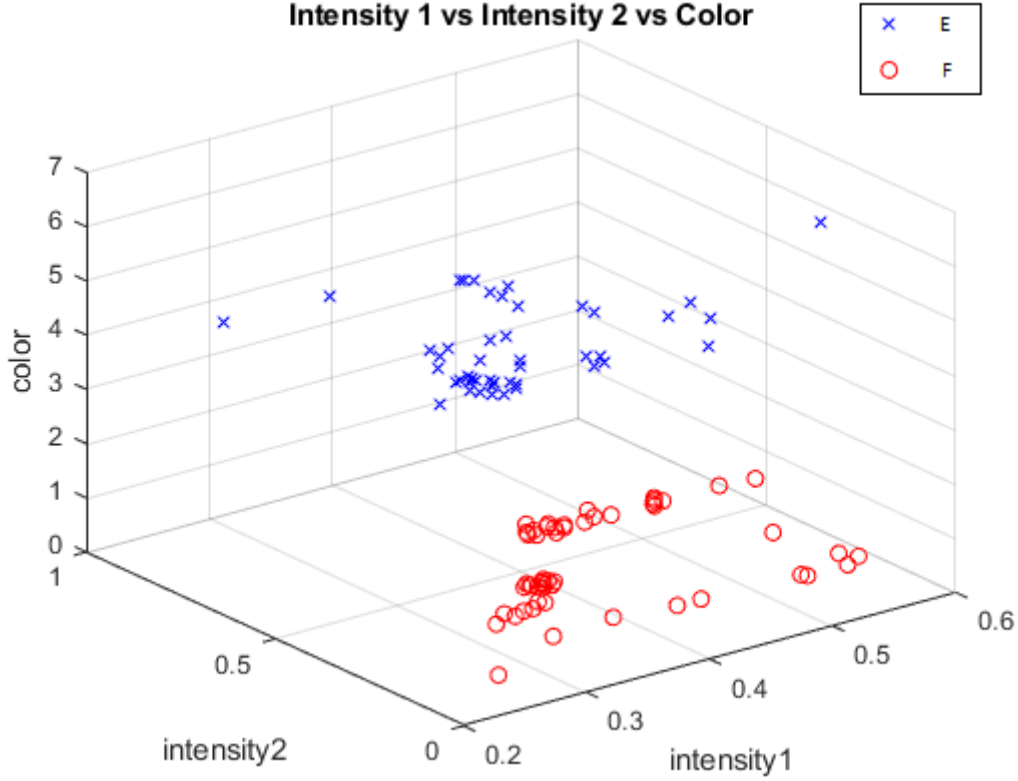
Figure 17 Plot Colour Cluster Scenario 2, Division 2. The intensity of sensor 1 is plotted on the y-axis (right axis), the intensity of sensor 2 is shown on the x-axis (left axis) and the colour ID is plotted on the z-axis (vertical axis). Cluster E contains the samples with high (colour) intensity with a blue 'x' and cluster F marks the trials with low intensity with a red dot.

Probabilities and Size of Prediction Error

Kullback-Leibler Divergence

The model is based on the chances that a data point is in a combined cluster. For scenario 1, these combined clusters are different than for scenario 2. In the first part of scenario 1, the combined clusters are (A,C), (A,D), (B,C), and (B,D). This categorization is based on the motor rotation clustering and the intensity clustering. The second part of scenario 1 is based on the categorization of intensity clustering and colour clustering. The combined clusters are (C,E), (C,F), (D,E), and (D,F). For scenario 2, the combined clusters are (A,E), (A,F), (B,E), and (B,F). These categorisations will result for every part of both scenarios in two probability distributions. These distributions are the generative model. The Kullback-Leibler divergence will be used to calculate the size of the prediction error.

$$D_{KL}(Pr_{obs}||Pr_{pred}) = \sum_{p \in \Omega(obs)} Pr_{obs}(p) \times log_2\left(\frac{Pr_{obs}(p)}{Pr_{pred}(p)}\right)$$

Table 3 and Table 4 show the structure of the following tables, with the expectation of the probabilities. The difference between the tables is that in the first tables, there is a column with

clusters, where in the following tables, the size of the prediction error will be showed in those columns. Every table will be followed by two plots: one plotting the probabilities and one plotting the sizes of the prediction error (PE).

Scenario 1		turn right		turn left	
		probability	clusters	probability	clusters
intensity	low	P = high	(A,C)	P = low	(B,C)
	high	P = low	(A,D)	P = high	(B,D)

		intensity			
		low		high	
		probability	clusters	probability	clusters
colour intensity	low	P = high	(C,E)	P = low	(D,E)
	high	P = low	(C,F)	P = high	(D,F)

Table 3 Structure of Probability Tables Scenario 1, given the motor movement of the robot and the measured intensity and colour. The expected probability and the cluster that value is in are given.

Scenario 2		turn right		turn left	
		probability	clusters	probability	clusters
colour intensity	low	P = high	(A,E)	P = low	(B,E)
	high	P = low	(A,F)	P = high	(B,F)

Table 4 Structure of Probability Tables Scenario 2. It shows the clusters in which a trial can be classified and the expected probability of being in those clusters.

Scenario 1 : first division

When turning right (cluster A), there are 50 trials in cluster C and no trials in cluster D. 32 trials are in cluster C, when turning left (cluster B) , which means that there are 18 trials in cluster D when turning left. In cluster C, there are 82 trials, of which 32 are in cluster E and 50 are in cluster F. The resulting trials are all 18 in cluster D and F. The probability distribution and the size of the prediction errors based on this classification are shown in Table 5 and plotted in Figure 19 and Figure 20.

Scenario 1		turn right		turn left	
		probability	prediction error	probability	prediction error
intensity	low	1	0	0.64	0.6439
	high	0	∞	0.36	1.4739
		intensity			
		low		high	
		probability	prediction error	probability	prediction error
colour intensity	low	0.3902	1.3577	0	∞
	high	0.6098	0.7136	1	0

Table 5 Probabilities & Size of PE Scenario 1, Division 1
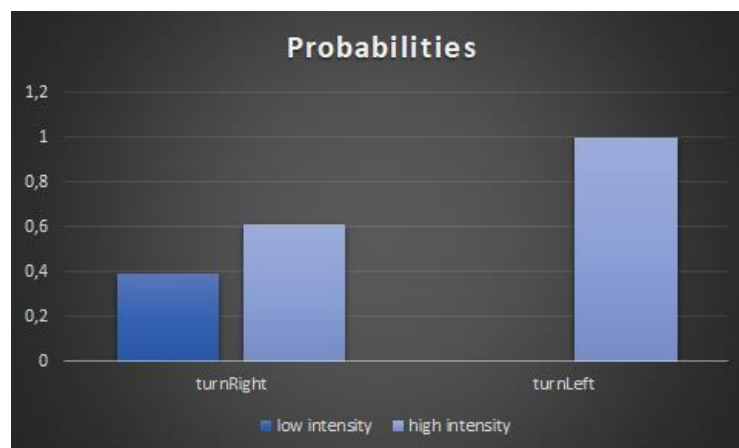
Figure 18 Probabilities Scenario 1, Division 1

Figure 19 Sizes of PE Scenario 1, Division 1

Scenario 2 : first division

In scenario 2, half of the trials where the robot is turning right (cluster A) are in cluster E and the other half is in cluster F. When turning left (cluster B), there are 7 trials in cluster E and 43 trials in cluster F. The probability distribution and the size of the prediction errors based on this classification are shown in Table 6 and plotted in Figure 21 and Figure 22.

Scenario 2		turn right		turn left	
		probability	prediction error	probability	prediction error
colour intensity	low	0.5	1	0.14	2.8365
	high	0.5	1	0.86	0.2176

Table 6 Probabilities & size PE Scenario 2, Division 1
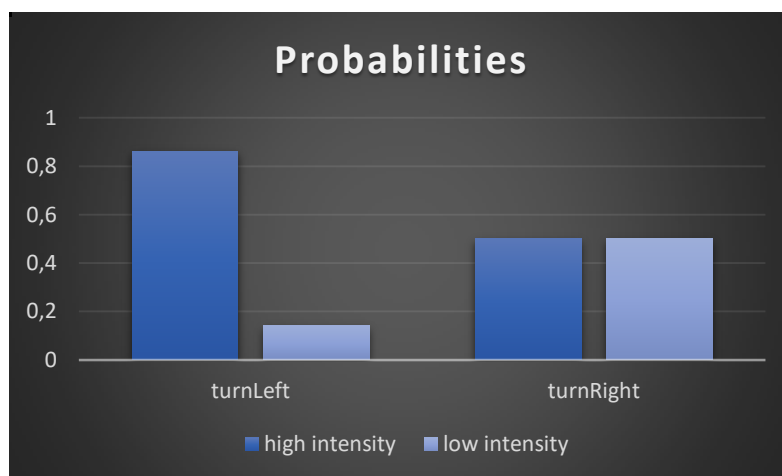
Figure 20 Probabilities Scenario 2, Division 1

25
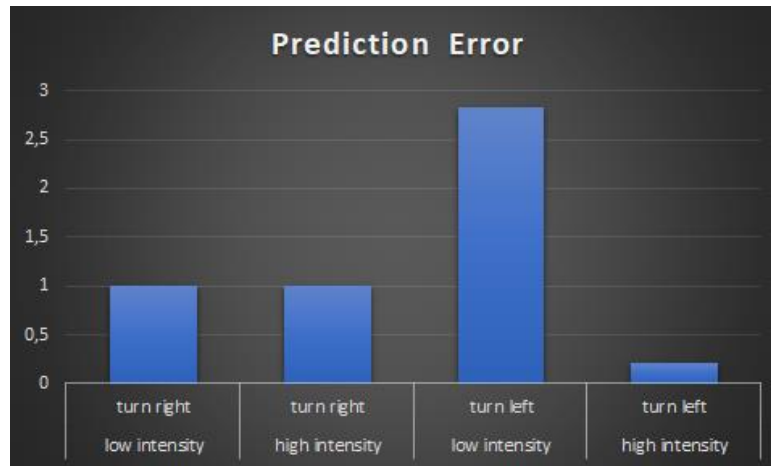
Figure 21 Sizes of PE Scenario 2, Division 1

Scenario 1: second division

For the first part, every trial that is divided in cluster A, also belongs to cluster C. When turning left (cluster B), 32 trials are clustered in cluster C, and 18 trials are in cluster D. In the second part, all the 18 trials of cluster D also belong to cluster F. 54 trials belong to cluster C and E, so 28 trials are in cluster C and F. The probability distribution and the size of the prediction errors based on this classification are shown in Table 7 and plotted in Figure 23 and Figure 24.

Scenario 1		turn right		turn left	
		probability	prediction error	probability	prediction error
intensity	low	1	0	0.64	0.6439
	high	0	∞	0.34	1.4739

		intensity			
		low		high	
		probability	prediction error	probability	prediction error
colour intensity	low	0.6585	0.6027	0	∞
	high	0.3415	1.55	1	0

26

Figure 22 Probabilities Scenario 1, Division 2
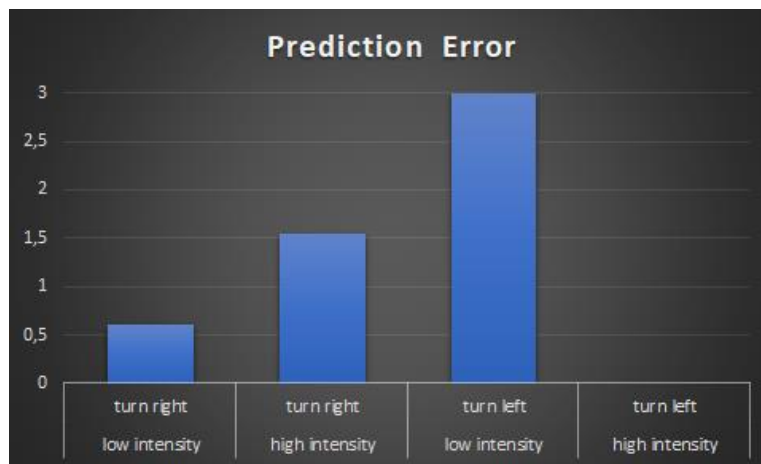
Figure 23 Sizes of PE Scenario 1, Division 2

Scenario 2 : second division

When turning right (cluster A), 47 trials are in cluster E, while 3 trials are in cluster F. For turning left (cluster B), 7 trials are in cluster E, and 43 trials are in cluster F. The probability distribution and the size of the prediction errors based on this classification are shown in Table 8 and plotted in Figure 25 and Figure 26.

Scenario 2		turn right		turn left	
		probability	prediction error	probability	prediction error
colour intensity	low	0.94	0.0893	0.14	2.8365
	high	0.06	4.0589	0.86	0.2176

Table 8 Probabilities & size of PE Scenario 2, Division 2
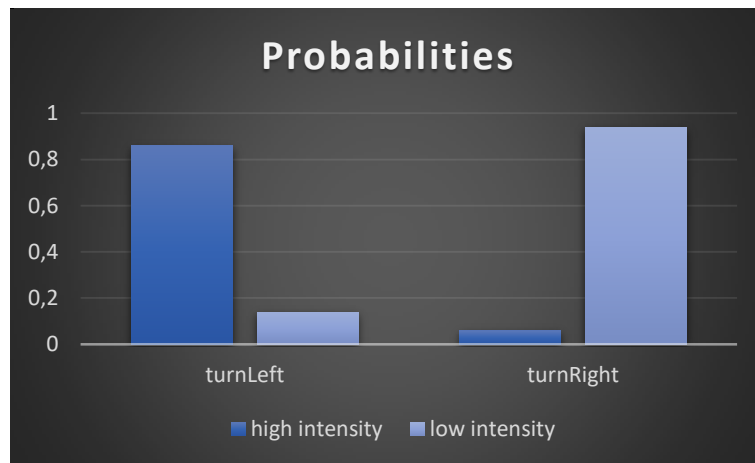
Figure 24 Probabilities Scenario 2, Division 2

Figure 25 Sizes of PE Scenario 2, Division 2

Total size of prediction error

The total size of prediction error per scenario and per division are computed below and plotted in Figure 27.

Scenario 1 & division 1 : $32 \times 1.3577 + 50 \times 0.716 = $ **79.1264**

Scenario 2 & division 1 : $25 \times 1 + 25 \times 1 + 7 \times 2.8365 + 43 \times 0.2176 = $ **79.2123**

Scenario 1 & division 2 : $54 \times 0.6027 + 28 \times 1.55 = $ **75.9458**

Scenario 2 & division 2 : $47 \times 0.0893 + 3 \times 4.0589 + 7 \times 2.8365 + 43 \times 0.2176 = $ **45.5861**
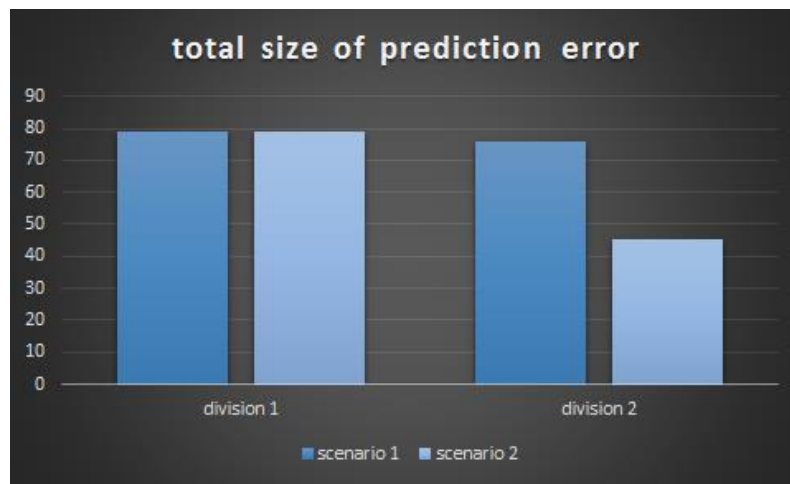
Figure 26 Total Size of Prediction Errors

Discussion

In this section, I will first give a short overview of my research question and the proposed hypotheses. Then, I will answer my research question for both the divisions I found. Afterwards, I will discuss possible explanations of these results, which I will relate to the empirical developmental literature. I conclude this section by giving some possible research questions for future research.

New-borns start without colour vision, meaning that they are only able to distinguish between brightness of colours. After fifteen days, they gradually start to learn different colours, showed by Adams and colleagues (1987), and Clavadetscher and colleagues (1988). Based on their research, I opted for two different scenarios to implement in my robot research. One scenario is that the robot step-wise learns to distinguish between intensity and colour. First it learns intensities based on motor movement, afterwards it adjusts the model to the colour perception. In the second scenario, the model immediately learns intensity and colour based on motor movement. The first scenario is based on how children's vision develops, so my expectation was that the total prediction error of this scenario would be lower than the total prediction error of scenario 2. Therefore, my research question is: "What is the difference in the size of the prediction error between the two different scenarios?". My results showed two divergent divisions, so I will answer the research question separately for both divisions.

Division 1

For division 1, there is no real difference in total size of the prediction error between the two distinct scenarios. The total size of the prediction error of scenario 1 is 79.1264, while the total size of the prediction error of scenario 2 is 79.2123. The difference between the sizes of the prediction error is 0.0859, which is very small. To answer my research question on this part of the results: there is no difference in the size of the prediction error between the different scenarios for division one.

Division 2

The total size of the prediction error in scenario 1 is 75.9458, while the total size of the prediction error in scenario 2 is 45.5861. The difference in size of prediction error between scenario 1 and scenario 2 is 30.3597, which is a lot higher than the difference in division 1. For this division scenario 2 gives the best model with the lowest size of prediction error.

There exist many possible explanations for the missing difference in division 1 and the difference contrary to my expectation in scenario 2. First of all, the measurements of the robot were not optimal. There were nine wrong colour measurements (green was not perceived correct for four times, red was not recognized correct in five trials). There were also unexpected values, but these cannot be proven to be caused by external circumstances, so I kept my data intact. If I deleted the outliers (based on the

outlier function in MATLAB), I would have cancelled 28 trials of the 100, resulting in unreliable remaining data. Another reason is that sensor 1 gives low intensity for bright colours, because it is not turned to the light. In this way, the value of intensity 1 and the value of intensity 2 are opposite, resulting in a potentially wrong cluster division.

The size of the prediction error depends on the clustering algorithm (because of this, the two divisions gave a different size of prediction error). Since the clustering algorithm can result in varying divisions per run and the probabilities and the prediction errors depend on the division, the total size of the prediction error varies per run. If the clustering algorithm resulted in a perfect division, there could possibly be a bigger difference, or even a difference in the opposite direction. Another possible reason is that the total amount of samples is small, just as the domain. The difference could possibly be bigger if additional information and an environment with more context were provided, because of a larger amount of samples could result in a better learned model.

To conclude, I did not find the difference that was expected based on the literature about colour vision development in infants, (Adams at al., 1987, Clavadetscher at al., 1988) for the reasons mentioned above. However, the question of how PE is dealt with during infancy is an important question in understanding the development of the generative models and hence needs to be further explored. A few possible future research questions are discussed next.

Infants in development continuously perceive new samples to classify, where a robot only samples discrete values in a limited range of time. The environment of a child is richer than the environment of my robot. Based on this, I think that the need for gradual development is necessary and can be proven by experimenting with a robot in a richer environment and more samples, mimicking the child development better. If the robot had the same amount of information and samples, the need for gradual development would possibly be higher than in this research.

The robot that I used only has two wheels, two sensors, and took 100 trials. Humans, however, have a much richer environment, with more senses and more samples. Because of this, the research could be expanded with a bigger environment, more sensors, and more samples. When using more actuators, what would this mean for the clustering algorithm? Is it necessary to take a different value for k, or is adding a dimension good enough? If the robot takes more samples, would it be better to update the clusters after each trial, to see if online updating results in better clusters? Besides this, the movement and placement of the robot were simplified in this study. It would be interesting to see how the model reacts when the robot is programmed to move in all directions, instead of only circular movement. How would variation in the starting position of the robot change the model? What effect would

starting the next trials where the previous one ended, or moving completely randomly have on the clusters?

Conclusion

New-borns start by perceiving only the intensity of objects and they gradually develop colour vision. I modelled this development by first learning intensities based on motor movement, and afterwards adding colour perception. A second scenario was modelled, by immediately learning intensities and colours based on motor movement. This resulted in my research question : "What is the difference in the size of the prediction error between step-wise learning intensity and colour perception based on movement (Scenario 1), and immediately learning intensity and colour perception based on motor movement (Scenario 2)?". I tested both scenarios and compared the scenarios based on the total size of prediction errors computed by using the Kullback-Leibler divergence. Different clustering resulted in two distinct divisions. For division one, there was no difference between total size of prediction errors. In the second division, the immediate learning of intensities and colours based on motor movement gave a lower error than the step-wise learning of intensities and colours respectively, based on motor movement.. I did not find the difference in total size of prediction error that we expected based on the literature about colour vision development in infants.

References

Adams, R. J. (1987). An evaluation of color preference in early infancy. *Infant Behavior and Development*, *10*(2), 143–150. http://doi.org/10.1016/0163-6383(87)90029-4

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., … Yoshida, C. (2009). Cognitive Developmental Robotics: A Survey. *IEEE Transactions on Autonomous Mental Development*, *1*(1), 12–34. http://doi.org/10.1109/TAMD.2009.2021702

Clark, A. (2013). Whatever next ? Predictive brains , situated agents , and the future of cognitive science, 1–73. http://doi.org/10.1017/S0140525X12000477

Clavadetscher, J. E., Brown, a M., Ankrum, C., & Teller, D. Y. (1988). Spectral sensitivity and chromatic discriminations in 3- and 7-week-old human infants. *Journal of the Optical Society of America. A, Optics and Image Science*, *5*(12), 2093–105. http://doi.org/10.1364/JOSAA.5.002093

Deák, G. O., Fasel, I., & Movellan, J. (2001). *The Emergence of Shared Attention : Using Robots to Test Developmental Theories*. San Diego, USA. Retrieved from https://pdfs.semanticscholar.org/066c/128c721594636e24a69ad30ce48880bc2fdf.pdf

Kumar.(2014). *Introduction to Data Mining*. Harlow: Pearson.

Kwisthout, J., Bekkering, H., & Rooij, I. Van. (2017). To be precise , the details don ' t matter : On predictive processing , precision , and level of detail of predictions. *Brain and Cognition*, *112*, 84–91. http://doi.org/10.1016/j.bandc.2016.02.008

Kwisthout, J. (2016). ERC Starting Grant 2016 Research proposal [Part B2]

K-means clustering. (n.d.). Retrieved May 1, 2018, from https://nl.mathworks.com/help/stats/kmeans.html

Otworowska, M., Riemens, J., Kamphuis, C., Wolfert, P., Vuurpijl, L., & Kwisthout, J. (2015). The Robo-havioral Methodology : Developing Neuroscience Theories with FOES. In *Proceedings of the 27th Benelux Conference on AI (BNAIC'15)* (p. 8). Nijmegen, The Netherlands. Retrieved from www.socsci.ru.nl/johank/RoboHavioral.pdf

Peeles, D., & Teller, D. Y. (1975). Color Vision and Brightness Discrimination in Two-Month-Old Human Infants. *Science*, *189*(4208), 1102–1103. Retrieved from http://science.sciencemag.org/content/189/4208/1102.long

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. http://doi.org/10.1016/j.cognition.2010.11.015

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, *5*(2), 97–118. http://doi.org/10.1080/17588928.2013.877880

Ter Borg, M. M. C. (2017). *How generative models are created in Predictive Processing*. Radboud University Nijmegen. Retrieved from http://theses.ubn.ru.nl/bitstream/handle/123456789/4377/Borg%2C ter M._BSc_Thesis_2017.pdf?sequence=1

Wilton, P. C. (1937). Color vision in infants. *Journal of Experimental Psychology*, *20*(3), 203–222. Retrieved from http://psycnet.apa.org/record/1937-02990-001

Appendix

Appendix 1 : robot data and clustering division 1

Degrees Left Wheel	Degrees Right Wheel	Colour ID	Intensity 1	Intensity 2	Motor Cluster	Intensity Cluster	Colour Cluster
-331	307	5	0,2899	0,26	1	1	2
-567	505	4	0,5379	0,45	1	2	2
-618	146	4	0,3035	0,4	1	2	2
-560	375	4	0,4315	0,43	1	2	2
-384	482	4	0,4334	0,39	1	2	2
-402	17	5	0,2811	0,19	1	1	2
-160	312	1	0,285	0,17	1	1	1
-371	420	5	0,3778	0,22	1	1	2
-243	444	1	0,3182	0,17	1	1	1
-454	72	5	0,2879	0,18	1	1	2
-481	299	5	0,364	0,19	1	1	2
-79	266	7	0,2723	0,09	1	1	2
-435	107	1	0,2879	0,15	1	1	1
-207	310	1	0,2977	0,16	1	1	1
-401	542	4	0,5067	0,35	1	2	2
-209	434	1	0,3156	0,18	1	1	1
-353	115	7	0,2791	0,23	1	1	2
-464	382	5	0,4256	0,42	1	2	2
-329	415	5	0,3407	0,39	1	2	2
-177	169	7	0,2713	0,24	1	1	2
-290	299	5	0,2987	0,41	1	2	2
-287	506	5	0,3534	0,39	1	2	2
-349	512	5	0,4256	0,39	1	2	2
-610	54	4	0,284	0,91	1	2	2
-184	278	5	0,283	0,34	1	2	2
-119	477	5	0,2889	0,27	1	1	2
-247	98	7	0,2742	0,24	1	1	2
-557	298	4	0,367	0,9	1	2	2
-118	482	5	0,2967	0,38	1	2	2
-316	359	5	0,3104	0,4	1	2	2
-421	501	5	0,4852	0,33	1	2	2
-300	86	7	0,2882	0,17	1	1	2
-430	43	5	0,2781	0,21	1	1	2
-256	112	7	0,2733	0,17	1	1	2
-506	96	5	0,2801	0,27	1	1	2
-331	42	5	0,2752	0,23	1	1	2
-436	232	5	0,2957	0,23	1	1	2
-199	530	5	0,3045	0,2	1	1	2
-217	498	5	0,3045	0,22	1	1	2

-618	243	4	0,3729	0,52	1	2	2
-241	200	5	0,2762	0,27	1	1	2
-138	456	5	0,2899	0,22	1	1	2
-614	528	4	0,5194	0,5	1	2	2
-343	296	5	0,3006	0,19	1	1	2
-210	524	1	0,3054	0,16	1	1	1
-79	598	5	0,2957	0,24	1	1	2
-147	318	7	0,2742	0,14	1	1	2
-63	282	1	0,2694	0,14	1	1	1
-54	471	5	0,284	0,26	1	1	2
-415	215	5	0,2899	0,28	1	1	2
119	-444	6	0,2733	0,09	2	1	2
385	-552	1	0,4803	0,02	2	1	1
471	-141	3	0,2814	0,02	2	1	2
24	-623	3	0,2723	0,05	2	1	2
309	-592	1	0,3827	0,03	2	1	1
38	-561	2	0,2703	0,05	2	1	1
338	-426	3	0,3182	0,04	2	1	2
547	-533	1	0,5302	0,03	2	1	1
610	-90	0	0,2742	0,15	2	1	1
199	-436	2	0,2694	0,06	2	1	1
391	-278	3	0,2908	0,03	2	1	2
155	-2	2	0,2811	0,03	2	1	1
363	-459	3	0,367	0,04	2	1	2
156	-435	3	0,2664	0,05	2	1	2
630	-223	3	0,3631	0,03	2	1	2
125	-8	2	0,2772	0,04	2	1	1
526	-266	3	0,329	0,03	2	1	2
131	-125	2	0,2694	0,05	2	1	1
478	-629	7	0,5214	0,1	2	1	2
475	-195	3	0,286	0,04	2	1	2
330	-323	3	0,283	0,05	2	1	2
285	-316	3	0,2869	0,06	2	1	2
495	-330	1	0,4022	0,03	2	1	1
346	-33	2	0,2645	0,05	2	1	1
69	-570	3	0,2752	0,08	2	1	2
380	-179	3	0,2645	0,04	2	1	2
426	-249	3	0,2938	0,04	2	1	2
46	-75	2	0,283	0,06	2	1	1
316	-507	1	0,3368	0,05	2	1	1
361	-412	3	0,3221	0,07	2	1	2
171	-356	6	0,2673	0,07	2	1	2
420	-317	3	0,3612	0,02	2	1	2
603	-618	1	0,5233	0,06	2	1	1
459	-89	3	0,2684	0,03	2	1	2

579	-338	1	0,4825	0,01	2	1	1
284	-204	2	0,2732	0,03	2	1	1
603	-289	3	0,4432	0,02	2	1	2
126	-568	1	0,285	0,04	2	1	1
191	-258	2	0,2704	0,05	2	1	1
443	-408	3	0,4168	0,03	2	1	2
89	-389	2	0,2791	0,05	2	1	1
328	-449	3	0,3709	0,03	2	1	2
297	-53	2	0,2723	0,04	2	1	1
11	-21	2	0,2859	0,04	2	1	1
619	-158	3	0,3074	0,03	2	1	2
177	-552	2	0,284	0,05	2	1	1
538	-485	1	0,5151	0,01	2	1	1
339	-548	3	0,37	0,05	2	1	2
535	-612	1	0,5128	0,2	2	1	1
167	-385	2	0,2791	0,05	2	1	1

Appendix 2 : robot data and clustering division 2

Degrees Left Wheel	Degrees Right Wheel	Colour ID	Intensity 1	Intensity 2	Motor Cluster	Intensity Cluster	Colour Cluster
-331	307	5	0,28986	0,26	2	2	2
-567	505	4	0,5379	0,45	2	1	2
-618	146	4	0,3035	0,4	2	1	2
-560	375	4	0,4315	0,43	2	1	2
-384	482	4	0,4334	0,39	2	1	2
-402	17	5	0,2811	0,19	2	2	2
-160	312	1	0,285	0,17	2	2	1
-371	420	5	0,3778	0,22	2	2	2
-243	444	1	0,3182	0,17	2	2	1
-454	72	5	0,2879	0,18	2	2	2
-481	299	5	0,364	0,19	2	2	2
-79	266	7	0,2723	0,09	2	2	2
-435	107	1	0,2879	0,15	2	2	1
-207	310	1	0,2977	0,16	2	2	1
-401	542	4	0,5067	0,35	2	1	2
-209	434	1	0,3156	0,18	2	2	1
-353	115	7	0,2791	0,23	2	2	2
-464	382	5	0,4256	0,42	2	1	2
-329	415	5	0,3407	0,39	2	1	2
-177	169	7	0,2713	0,24	2	2	2
-290	299	5	0,2987	0,41	2	1	2
-287	506	5	0,3534	0,39	2	1	2
-349	512	5	0,4256	0,39	2	1	2

-610	54	4	0,284	0,91	2	1	2
-184	278	5	0,283	0,34	2	1	2
-119	477	5	0,2889	0,27	2	2	2
-247	98	7	0,2742	0,24	2	2	2
-557	298	4	0,367	0,9	2	1	2
-118	482	5	0,2967	0,38	2	1	2
-316	359	5	0,3104	0,4	2	1	2
-421	501	5	0,4852	0,33	2	1	2
-300	86	7	0,2882	0,17	2	2	2
-430	43	5	0,2781	0,21	2	2	2
-256	112	7	0,2733	0,17	2	2	2
-506	96	5	0,2801	0,27	2	2	2
-331	42	5	0,2752	0,23	2	2	2
-436	232	5	0,2957	0,23	2	2	2
-199	530	5	0,3045	0,2	2	2	2
-217	498	5	0,3045	0,22	2	2	2
-618	243	4	0,3729	0,52	2	1	2
-241	200	5	0,2762	0,27	2	2	2
-138	456	5	0,2899	0,22	2	2	2
-614	528	4	0,5194	0,5	2	1	2
-343	296	5	0,3006	0,19	2	2	2
-210	524	1	0,3054	0,16	2	2	1
-79	598	5	0,2957	0,24	2	2	2
-147	318	7	0,2742	0,14	2	2	2
-63	282	1	0,2694	0,14	2	2	1
-54	471	5	0,284	0,26	2	2	2
-415	215	5	0,2899	0,28	2	2	2
119	-444	6	0,2733	0,09	1	2	2
385	-552	1	0,4803	0,02	1	2	1
471	-141	3	0,2814	0,02	1	2	1
24	-623	3	0,2723	0,05	1	2	1
309	-592	1	0,3827	0,03	1	2	1
38	-561	2	0,2703	0,05	1	2	1
338	-426	3	0,3182	0,04	1	2	1
547	-533	1	0,5302	0,03	1	2	1
610	-90	0	0,2742	0,15	1	2	1
199	-436	2	0,2694	0,06	1	2	1
391	-278	3	0,2908	0,03	1	2	1
155	-2	2	0,2811	0,03	1	2	1
363	-459	3	0,367	0,04	1	2	1
156	-435	3	0,2664	0,05	1	2	1
630	-223	3	0,3631	0,03	1	2	1
125	-8	2	0,2772	0,04	1	2	1
526	-266	3	0,329	0,03	1	2	1
131	-125	2	0,2694	0,05	1	2	1

478	-629	7	0,5214	0,1	1	2	2
475	-195	3	0,286	0,04	1	2	1
330	-323	3	0,283	0,05	1	2	1
285	-316	3	0,2869	0,06	1	2	1
495	-330	1	0,4022	0,03	1	2	1
346	-33	2	0,2645	0,05	1	2	1
69	-570	3	0,2752	0,08	1	2	1
380	-179	3	0,2645	0,04	1	2	1
426	-249	3	0,2938	0,04	1	2	1
46	-75	2	0,283	0,06	1	2	1
316	-507	1	0,3368	0,05	1	2	1
361	-412	3	0,3221	0,07	1	2	1
171	-356	6	0,2673	0,07	1	2	2
420	-317	3	0,3612	0,02	1	2	1
603	-618	1	0,5233	0,06	1	2	1
459	-89	3	0,2684	0,03	1	2	1
579	-338	1	0,4825	0,01	1	2	1
284	-204	2	0,2732	0,03	1	2	1
603	-289	3	0,4432	0,02	1	2	1
126	-568	1	0,285	0,04	1	2	1
191	-258	2	0,2704	0,05	1	2	1
443	-408	3	0,4168	0,03	1	2	1
89	-389	2	0,2791	0,05	1	2	1
328	-449	3	0,3709	0,03	1	2	1
297	-53	2	0,2723	0,04	1	2	1
11	-21	2	0,2859	0,04	1	2	1
619	-158	3	0,3074	0,03	1	2	1
177	-552	2	0,284	0,05	1	2	1
538	-485	1	0,5151	0,01	1	2	1
339	-548	3	0,37	0,05	1	2	1
535	-612	1	0,5128	0,2	1	2	1
167	-385	2	0,2791	0,05	1	2	1

Appendix 3 : group proposal

Group Project Proposal for BSc AI Thesis Research "PREDICT"

Group constellation

Supervisor: Johan Kwisthout, Danaja Rutar
Students: Bea Waelbers, Jet van Dijk, Borislav Sabev, Casper van Aarle

Project description

1. Group project title
PREDICT: Predictive Robots Empirically Determine Influential Cognitive Theories

2. Abstract: (max 100 words) *include a word count: 100*

Predictive Processing claims to be a unifying account that describes all of cognition. However, the account has been fleshed out only at the level of low-level perception. For higher cognition, e.g., action understanding, communication, and problem solving, it is sketchy at the best. In our experience, theoretical gaps, under-defined concepts, and ambiguities in a verbal theory become manifest when explicating the theory into computational models and implement them. We propose to implement and test parts of the predictive processing principle, partially using experiments with Mindstorms robots. We focus on conceptual and computational aspects of the development of generative models in PP.

3. Brief Project description: (max. 400 words)

Background and motivation

The predictive processing account has gained considerable interest in contemporary cognitive neuroscience. Its key idea is that the brain is in essence a hierarchically organized hypothesis-testing mechanism, continuously attempting to minimize the error of its predictions. More precisely, the account assumes a hierarchy of increasingly abstract (probabilistic) predictions, and the hypothesized causes that drive the predictions. At each level of the hierarchy, the predictions about the inputs are compared with the actual inputs, and possible prediction errors are minimized. A crucial aspect of the theory – how are the generative models actually developed and how are they shaped by prediction errors– is yet overlooked. In this project we focus on various aspects of this foundational issue.

Main aims and research questions of the project

We aim to study an important open problem in the predictive processing account (how are generative models developed and revised) by means of conceptual analysis, computational and formal modeling, and robot experimentation and exploration. Here we build on previous bachelor projects within this theme, in particular the projects of Maaike ten Borg, Djamari Oetringer, Erwin de Wolff and Dennis Verheijden. In this group project we study:

a. How developmental aspects, such as the emergence of color vision and the integration of auditory and visual cues, can affect the learning of generative models;
b. How exploration strategies can be guided by characteristics of the generative model;
c. How the trade-off between level-of-detail and prediction error can be resolved for a specific action requirement.

Research plan (approach, methods, design, analyses)

The research approach consists of conceptual analysis, computational and formal modeling, robot-construction and exploration, and programming, in particular further development of the Predictive Processing Toolbox. **Bea and Jet** will focus on research question a, **Caspar** on research question b, and **Borislav** on research question c.

The group members are invited to attend the Predictive Processing Research Seminar (http://www.socsci.ru.nl/johank/seminar.html) and the Donders Predictive Processing PI group at a regular basis.

4. Schedule: (max. 1 page)

See individual projects.

5. Backup plan:

Name the backup supervisor who takes over in emergency cases

Johan and Danaja are each other's backup.

6. Main References: (max. 10)

Johan Kwisthout, Harold Bekkering, and Iris van Rooij (2017). To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions. Brain and Cognition, 112 (special issue Perspectives on Human Probabilistic Inference), 84-91.

Johan Kwisthout and Iris van Rooij (2015). Free energy minimization and information gain: The devil is in the details. Commentary on Friston, K., Rigoli, F., Ognibene, D., Mathys, C., FitzGerald, T., and Pezzulo, G. (2015). Active Inference and epistemic value. *Cognitive Neuroscience, 6*(4), 216-218.

Maria Otworowska, Jordi Riemens, Chris Kamphuis, Pieter Wolfert, Louis Vuurpijl, and Johan Kwisthout (2015). *The Robo-havioral Methodology: Developing Neuroscience Theories with FOES*. Proceedings of the 27th Benelux Conference on AI (BNAIC'15), November 5-6, Hasselt, Belgium.