PROBABILISTIC TOOLS TO GAUGE 6-MONTH INFANTS' ABILITY TO DETECT CONTINGENCY

Alice Dauphin supervised by J. Kevin O'Regan

Masters thesis presented to M2 Neurosciences, Comportement, Cognition

Université Paul Sabatier

31 May 2019

Abstract

Adults are capable of very fine motor skills whereas newborn babies' motions are less accurately adjusted to the environment. It has been suggested that babies are sensitive to sensorimotor contingencies so they can acquire their body know-how by gradually linking each body movement to its perceptual consequences. The research we pursued in the team is part of this theoretical framework. We use behavioural measurements to study how babies refine their body know-how over time.

During my internship, we studied arm differentiation in infants of age 6 months. An artificial contingency was established between the movements of one of the babies' arms and the appearance of visual and auditory stimuli on both of their arms. My goal was to develop analytical tools to assess if babies detect the contingency (i.e. if they realize that they caused the occurrence of the stimuli). I tried to reproduce the probabilistic methodology developed by J. Watson in his experiments with 4-month old babies. I could not obtain reliable results and so pursued my investigations. I adapted Watson's analytical tools to create a binary indicator measuring the success of babies at the individual level. I showed that babies can differentiate between a situation where without doubt they have no control and a situation where they could be the cause of the stimulus. However, because babies who tried to test the contingency behaved similarly in both the test and the control group I can not ascertain that babies from the contingent group understood that they triggered the contingency.

Résumé

Chez l'adulte, les mouvements sont ajustés très finement à l'environnement. Les nouveaux nés, en revanche, ont des mouvements moins bien maîtrisés c'est-à-dire, mal calibrés à leur environnement. Un des mécanismes proposé pour permettre l'acquisition du savoir-faire moteur postule qu'il existe, chez l'enfant, une sensibilité accrue aux contingences sensorimotrices. Petit à petit, les conséquences perceptuelles d'un geste seraient associés à ce même geste jusqu'à l'obtention d'un modèle corporel robuste. Les travaux de notre équipe s'insèrent dans ce cadre théorique. Nous étudions l'affinement du savoir-faire corporel au cours du développement à partir de mesures comportementales.

Pendant mon stage, je me suis intéressée aux indicateurs utilisés pour évaluer les performances des sujets. Nous avons testé la capacité d'enfants de 6 mois à différencier leurs bras en créant une contingence artificielle entre les mouvements de l'un des bras seulement et l'apparition de stimuli visuels et sonores sur les deux bras. J'ai essayé de mesurer la capacité des sujets à détecter la contingence artificielle via les outils d'analyse probabilistes développés par J. Watson dans ses expériences. Je n'ai pas obtenu de résultats fiables en traitant les données collectivement, et j'ai donc adapté ces outils pour les utiliser individuellement. J'ai montré que les enfants de 6 mois sont capables de distinguer une situation ambigüe où ils pourraient être à l'origine des stimuli d'une situation dans laquelle ils n'ont, sans aucun doute, pas d'impact sur leurs apparitions (bien trop nombreuses). Cependant, leurs comportements d'exploration étant similaires dans les deux situations, je ne peux pas, sans mesures supplémentaires, affirmer que les enfants du groupe contingent ont détecté la contingence même si plusieurs indices le suggèrent.

1. Introduction

The present work is being done in the team of Kevin O'Regan in the Integrative Neuroscience and Cognition Center (INCC, UMR 8002). It is part of the FEEL ERC Advanced Project Work Package 5 (WP5): Infant development and Robotics, and the FETopen project GoalRobots. The team is working on the acquisition of body know-how in babies and robots. Body know-how or body knowledge refers to an agent's ability to effectively control its body in its environment. Both in developmental psychology (e.g. Corbetta, 2018) and in the robotics literature (e.g. Oudeyer, Kaplan, & Hafner, 2007) it is generally accepted that agents must acquire their body know-how through their ability to detect **sensorimotor contingencies**, that is, their ability to detect the systematic co-occurrences (within a short time interval) of a **body movement (B)** with a **sensory stimulus (S)** (Richters & Eskew, 2009). Because usually there is a causal relationship between B and S, we shall call S the **responsive stimulus**. In psychology there is an extensive literature studying sensorimotor contingencies empirically. Much of this work involves experiments using variants of the so-called "mobile paradigm" developed by Rovee-Collier and her collaborators (Rovee & Rovee, 1969). In this paradigm, one of the baby's limbs is connected via a ribbon to a mobile placed above the cradle. When the infant moves, it produces auditory and visual feedback from the mobile. The set-up measures the baby's limb motions. This or similar methods have been used not only to study acquisition of body know-how, but also more generally to study memory, learning or motor development in babies (Jacquey, Fagard, Esseily, & O'Regan, in preparation). Underlying all this work is the global assumption that, within a few minutes of exposure, infants can relatively easily learn sensorimotor contingencies.

Three years ago, inspired by this literature, Kevin O'Regan's team started trying to replicate the classic results. They used a more "hi-tech" version of the mobile paradigm that uses bluetooth- or wifi-enabled accelerometers mounted in bracelets worn on the babies' limbs. These accelerometers made it possible to easily record the babies' motions. Each of the bracelets could also emit music, flash a light attached to it, or create a distal visual or auditory stimulation on a video monitor. Although the team performed one experiment with as many as 124 babies (Jacquey et al., submitted) and numerous additional pilot experiments (207 babies), curiously, the results have been weaker than expected, whatever the responsive stimulus chosen (auditory, visual, proximal, distal) and whatever the age of the babies (from 3 months to 9 months). It has become apparent that a possibility to explain this weakness is that the team has not been using the learning criterion used classically in the developmental literature.

In the literature, the criterion generally used was proposed in a founding article written by Fagen and collaborators (Fagen, Morrongiello, Rovee-Collier, & Gekoski, 1984). It is based on the activity of the relevant body part (in the mobile paradigm, it is the baby's limb): $\frac{\textit{time during which the limb is active}}{\textit{total duration of the experiment}} \times 100$. The word "active" refers to any measure of the movement indicator (acceleration, speed, strength…) above a chosen threshold. To judge whether a baby is sensitive to the contingency, its activity in a contingent condition is compared to its activity during a baseline period where there is no contingency. This is the learning criterion that is used classically:

$$\%\textit{activity in the contingent case} \geq \%\textit{activity during the baseline} \times 1{,}5$$

A first criticism of this way of testing whether babies have learnt the contingency is the arbitrariness of the criterion of 1.5. There would seem to be no mathematical justification for choosing this value, and it seems that researchers simply used it because previous researchers had used it. A more important criticism, and one that can explain why so many papers have found positive results, is related to the fact that babies tend generally to become more fussy over the course of an experiment. A baby's activity later in an experiment will therefore very often be greater than its activity in an initial period. If this initial period is used as a baseline, the criterion of 1.5 will often be exceeded, and experimenters will conclude that the baby has learnt the contingency, when in fact it has just become more active. This objection can be obviated by using a separate control condition with no contingency over the whole period of an experiment, yet in most of the later experiments using the mobile and related paradigms, such a separate control condition was not used. Indeed, recently, some workers have returned to the original use of a separate control group (Angulo-Barroso et al., 2017). Just as our team over the last three years, these workers have also shown less convincing results than previous literature suggested. It has even occasionally been observed that according to the criterion used classically, babies in the noncontingent groups actually could be claimed to have learnt a contingency when there was none present (Tiernan & Angulo-Barroso, 2008).

Should we conclude from all these observations that babies are actually much less sensitive to sensorimotor contingencies than has previously been thought? An alternative hypothesis is that babies are actually extremely sensitive to sensorimotor contingencies, but that they are able to learn them so quickly that they rapidly tire of exploring them. This would predict that individual babies might, during the course of an experiment, briefly explore a contingency, recognize it, and then immediately tire of it and pass to exploring some other aspect of the set-up. The reason previous work has only shown weak effects might therefore be that the effects are fleeting and occur at different moments in an experiment for different babies. Given this idea, I was asked to look in the literature to ascertain whether there existed other, more sensitive, moment-to-moment ways of demonstrating sensitivity to a contingency over the course of an experiment.

A search in literature led me to two series of articles written by Watson and by Butko & Movellan. These papers tackle the issue of contingency learning with a fruitful quantitative approach. Watson's papers (Watson, 1979, 1984, 1985) were aimed at building a theoretical framework to analyze and quantify contingency perception. Butko & Movellan (Movellan, 2005; Butko & Movellan, 2010) later made use of this material to suggest a strategy that could be used by a simple agent to detect (social) contingencies as quickly and accurately as possible. I used these articles as a starting point for the present work. Section 2 of this document thus provides a presentation of the methods developed by Watson. In Section 3 I then present an experiment I conducted with 44 6-month old babies in collaboration with Kevin O'Regan's team, in which I applied Watson's tools (Section 4). As we will see, a direct use of Watson's tools in our experiment is questionable, leading me to pursue my investigations in Section 5 where I developed another tool adapted from those of Watson, and applied it to our data (second analysis).

2. Watson’s criterion and results

The tools conceived by Watson to study contingency detection in babies are described in a number of different articles (Watson, 1979, 1984, 1985). The nomenclature and approaches in these articles are slightly different, but ultimately I decided to concentrate on his latest approach, summed up in a chapter he wrote in a book dedicated to the development of memory (Watson, 1984)[1]. Note that the links he made with memory in this chapter are not of concern here, and I have focused only on his probabilistic approach to contingency detection.

In order to study contingency perception, Watson’s team modified the mobile paradigm and applied it in an experiment with 146 babies of age 16 weeks. A color tv-screen showing a woman’s face was coupled to a pressure sensitive pillow on which lay the baby’s right foot. At the beginning of the experiment (1-minute *baseline*), the screen displayed the motionless face. After the baseline period, in the *contingent* group, a coupling was established whereby the woman’s face silently smiled or the mouth moved (responsive stimulus[2], S) when the infant pushed the pillow with its right leg (body movement, B). In the *non-contingent* group, the responsive stimulus (S) was delivered for 10% of the right-foot movements and randomly otherwise. The contingent or non-contingent sessions lasted 4 minutes (period of test).

The indicators used by Watson were: (1) The chance expectancy of observing B in a time span t following a randomly specified point in time (written P(B|random t)). It was calculated from the observed rate of B during the baseline and used for comparisons with (2), (3) and (4). (2) The probability of occurrence of B in a time range t following a prior occurrence of B (written P(B|Bt)). It was the reaction to the body movement itself (assessment of repetitive stereotypical gestures for example) observed in the non-contingent group during the period of test but also during the baseline. (3) The probability of occurrence of B in a time range t following S (written P(B|St)). It was the reaction to the stimulus observed in the

---

[1] For reference, a link to the pdf of this paper is available in the bibliography.

[2] I use the term “responsive stimulus” to refer to the result a body movement causes. Watson uses the term “stimulus” only and refers to the body movement as the “response”. I changed it to avoid confusion.

non-contingent group during the period of test. (4) The probability of occurrence of B in a time range t following a B→S conjunction (written P(B|B-St)). It was the reaction to a B→S conjunction observed in the contingent group during the period of test.

To do a time-dependent analysis, Watson had to define a specific span of time t wherein B was expected. If too long, t was irrelevant as it encompassed B for sure. If too short, t was irrelevant as it excluded B for sure. Watson set t at 3 seconds (the experiment lasted about 5 minutes). Notice that **the window of time t could either directly follow B or be delayed** (for example one might assume that there was a minimum latency below which a baby cannot respond) .

The probability of B occurring non-contingently (1) was calculated from the observed rate of B during the baseline: $\lambda = \frac{\text{number of B observed...}}{\text{...during the } 1-\text{minute baseline}}$. Its distribution was assumed to be the distribution of events occurring randomly at a constant average rate λ, i.e. a Poisson distribution[3]:

$$P(k \text{ events in interval } t) = \frac{(\lambda t)^k}{k!} e^{\lambda t}$$

$$P(k \geq 1) = \overline{P(0)} = 1 - P(0) = 1 - \frac{(\lambda t)^0}{0!} e^{\lambda t} = 1 - e^{\lambda t}$$

Then, P(B random) was compared with P(B|Bt), P(B|St), P(B|B-St). For each of these, the average difference with P(B random) and the standard error of the mean were calculated at 14 different moments following the occurrence of B, S or B→S. The mean of the standard errors over the 14 moments was calculated for each condition (dotted lines for each graph in figure 1 represent plus and minus twice the mean of the 14 standard errors).

This analysis of the subject's reactions to its own body movement, to the responsive stimulus, and to the contingency showed: (a) An excitatory reaction to the body movement itself in the first 2 minutes of the testing period in the non-contingent situation (i.e. a body movement immediately elicits another movement, see circle marked (a) in figure 1A) but no reaction to the body movement itself during the baseline (not shown). (b) An inhibitory reaction to the responsive stimulus during the four minutes of the testing period (first and last 2 minutes of

---

[3] We shall comment on this assumption later.

testing period were analyzed separately) in the non-contingent situation (i.e. a responsive stimulus inhibits body movements for 3 seconds, see circles marked (b) in figures 1C-1D). (c) An excitatory reaction to the B→S conjunction during the four minutes of the testing period in the contingent situation (a B→S conjunction elicits another body movement after about 7 seconds, see circles marked (c) in figures 1E-1F-1G). Notice that, this reaction to the B→S conjunction is added to the inhibitory reactions to B and S (marked (a+b) in figures 1E-1F-1G).

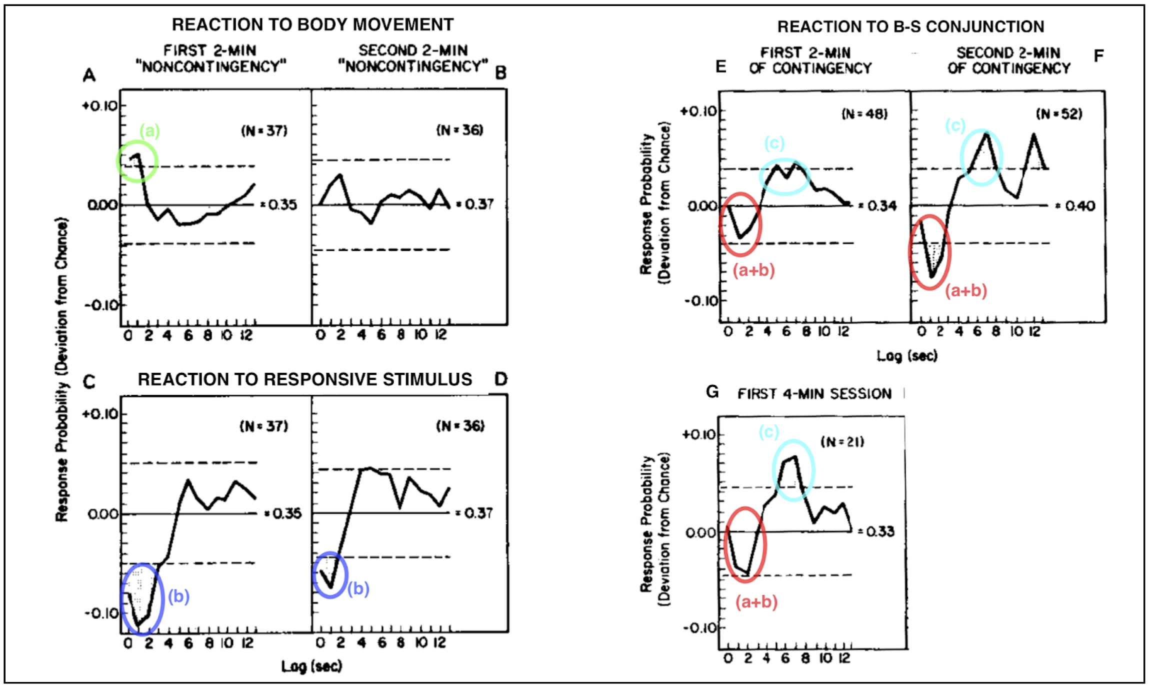

Figure 1 - Figures taken and annotated from Watson 1984. Means of the deviation from chance of P(B|Bt) (A-B), P(B|St) (C-D) and P(B|B-St) (E-F-G) during the first 2 minutes (A-C-E), from 2 minutes to 4 (B-D-F) or during the 4 minutes (G). The abscissae represent the lag after an initial body movement, responsive stimulus or B→S conjunction and the beginning of the 3-second window of time. The dotted lines represent plus and minus twice the mean of the 14 standard errors.

Watson assumed these results showed that the babies were sensitive to the contingency since a specific behavioral pattern, not expected from chance, occurred in the presence of a contingency, and since the reaction following a B→S conjunction could not be explained simply by the combined effects observed separately when B and S occurred. This would have been the case if the sum of the graphs A and C (see figure 1 above) had produced the graph E. This is not the case as a new pattern (c) appears.

Watson's approach is an example of an objective behavioral measure of contingency learning. The behavioral pattern observed in the contingent group differs enough from chance to assert a learning of the contingency. **In our replication, presented next, of the experiments led by Lisa Jacquey in our team, will we also find a behavioral pattern similar to the one Watson found? Alternatively, will we find some other distinctive pattern that is peculiar to the contingent group?** In the first analysis to be presented below I apply the tools suggested by Watson and collaborators to our experiment.

3. My replication experiment

Given the difficulty that Kevin O'Regan's team has had in the past finding reliable evidence for sensitivity to sensorimotor contingencies in young babies, the team was looking for new conditions where such a sensitivity could be demonstrated easily. One modification that the team decided to make as compared to previous experiments was to make the responsive stimulus caused by the baby's movements be *proximal,* that is, to occur on the body part that moved. In most previous experiments, stimulation had been presented *distally,* that is, on a video screen and loudspeaker in front of the baby. In addition, the team decided to make use of the finer probabilistic tools to estimate contingency sensitivity from moment to moment developed by Watson, and that I have been analysing. Given the results obtained in the study led by Lisa Jacquey with 4, 6 and 8 months old babies we pursued investigation with 6 months old babies, where the results seemed strongest. The experiment was led by Sergiu Popescu, and done in collaboration with Lisa Jacquey, Kevin O'Regan, and Judith Vergne. My own role, in addition to doing the mathematical analysis of the data generated, was to help with settling the babies, and setting up and monitoring the data recording during the experimental session, alternating with Sergiu Popescu.

The experiment proceedings were the following. The babies sat on their caretaker's lap. They had bracelets on their wrists containing wifi-connected micro-controllers. These micro-controllers recorded the accelerations of each arm and one of them (called the master) activated the responsive stimulus (music and red light) on the two limbs (see Appendix B for more about the responsive stimulus). This activation occurred when the master registered an acceleration greater than a fixed threshold in the contingent group (see Appendix A for more about the experimental set-up). It occurred randomly in the control group. The stimulation occurred on both bracelets even if it was provoked by one arm only. We chose this set-up because we wanted to avoid the possibility that babies moved the arm with the master micro-controller simply because they were excited by the responsive stimulus on this arm rather than detecting the contingency.

As in previous experiments, the question posed here was: **are 6-month-old infants able to detect the contingency (i.e. do the babies in the test group realize that the action of one of their arms determines the music and flashing lights)? Will they realize that this contingency is determined by only one of their arms, namely the "master"?**

Methods

44 6-month-old babies born normally were tested (aged from 5.5 to 6.5 months). The families came voluntarily and gave their written informed consent to participate. They were recruited locally (Paris region) and were middle- to upper-middle class families. There were two groups: a contingent and a non-contingent (control) group. In the contingent group, babies experienced perfect and immediate contingency enabled by the motions of the master arm. The experimental session started with a demonstration: one occurrence of the responsive stimulus was automatically shown to the babies. After this occurrence, the experiment began with a low threshold of activation. Each time the baby activated the bracelets, the threshold was stepped up (shaping). After 5 steps, the baby reached a fixed threshold[4] shared by all babies and the testing period of 5 minutes started. At the end of the 5-minute testing period, there were 30 seconds (baseline) without stimulation (the bracelets were off). The babies who did not complete the 5-minute testing period were excluded. We had difficulty in fulfilling this condition, which is why, unlike Watson, we did not include an initial baseline period (bracelets off during one minute before the first appearance of the responsive stimulus). This difference with Watson's procedure has the consequence that our analysis will have to be somewhat different from Watson's (see below).

---

[4] The threshold was chosen during the pilot experiments. It is a trade-off between preventing babies from provoking the stimuli continually (assessment of the performance impossible) and discouraging them from playing with the bracelets (if nothing happens after the occurrence of the demonstrator stimulus they have no reason to be interested in the bracelets).

In the control group, babies were exposed to random occurrences of the responsive stimulus for five minutes. At the end of the 5-minute period, there were 30 seconds without stimulation. To expose infants of both contingent and control groups to a comparable number of responsive stimuli, we looked at the fraction of time infants from the contingent group saw the responsive stimulus. The distribution was bimodal with peaks of the value of this fraction at 0.11 and 0.29. We divided the control group into two subgroups according to these values. One subgroup was exposed to the responsive stimulus for a tenth and the other for a third of the 5-minute testing period. The same criteria of inclusion were used for both groups so babies who did not complete the 5-minute period in the control group were excluded.

The contingent group contained 31 babies (16 girls and 15 boys) and we analyzed the data of 18 of them (7 girls and 11 boys). 3 infants had technical problems (no data recorded or dead battery), 2 infants had low data quality, 7 infants displayed fussiness before the end of the testing period and one infant did not trigger the contingency at least 5 times during the 5-minute period of test (see Appendix C for more about the inclusion criteria). At the time of the present writing, the experiment was not finished so the non-contingent group contained only 13 babies (6 girls and 7 boys) of which we analyzed the data of 11 of them (6 girls and 5 boys) - one infant displayed fussiness before the end of the testing period and one infant did not trigger the contingency at least 5 times during the 5-minute period of test (see Appendix C for more about the inclusion criteria).

4. First analysis

Watson used the tools he developed to analyse data at the group level. Consequently, they require a good number of babies to be used. Since my internship ended before data from enough babies had been collected in the control group, I applied Watson's analysis to the contingent group only (but see Section 5 for a more limited analysis that includes the control group).

Application of Watson’s analysis to the contingent group

In our experiment the stimuli were caused by the accelerations above a fixed threshold of the master bracelet only (see Appendix A). Consequently, **in the following, B refers to “a body movement of the master bracelet where the acceleration is superior to the threshold”.**

First, I conducted exactly the same analyses as Watson’s team did. I split my data covering the 5-minute testing period into two 2-minute periods (the first 2 minutes and from 2 minutes to 4) and the span t considered was 3 seconds. As Watson did, I applied an arbitrary criterion of at least 5 events (B) per sampling period to enter into the analysis in order to be able to calculate P(B|B-St). With the babies who completed the 5-minute testing period I carried out an adaptation of Watson’s tools using the two following indicators:

(1) The chance expectancy of observing B in a time span t following a randomly specified point in time (written P(B|random t)). Unfortunately, since we had made the choice (see above) of not including a baseline period, P(B|random t) could not be calculated in the way Watson did, namely from the observed rate of B before any occurence of the responsive stimulus. Instead a reasonable way of estimating chance expectancy seemed to us to be to calculate it from the observed rate of B during the whole 5-minute period of test: $\lambda = \frac{number\ of\ B\ observed...}{...over\ the\ 5{-}minute\ testing\ period} \Rightarrow \ P(B|random\ t) = 1 - e^{\lambda t}$. However, this choice may be problematic, and we will reconsider it later.

(2) The probability of occurrence of B in a time range t following a B→S conjunction (written P(B|B-St)). To calculate P(B|B-St) I divided the number of windows of time t in which at least one occurrence of B happened following a B→S conjunction, by the number of windows of time t in which it could have happened, that is the total number of windows of time t following a B→S conjunction. I defined the conditional probability of X given Y in a bin of time t (written P(X|Yt)) as Watson did. Nevertheless, I am not completely certain that he defined X as “at least one occurrence of B”. I interpreted his article in this way to be able to compare P(B|B-St) to P(B|random t). This seemed reasonable since in a previous article (Watson, 1979) Watson defined t as “the interval width from any random point in time” and P(B|random t) as “the probability of *an* event’s occurrence within that interval”. That is why I concluded that X must be defined as “an occurrence of X” that is “at least one occurrence of X”.

For each baby, I calculated the difference between P(B|B-St) and P(B|random t) at different times following the occurrence of B→S (from 0 through 13 seconds). Then, I computed the mean difference for the 14 lags as well as the standard error of the mean.

Results

The two following graphs cover the first 2 minutes (left) and from 2 minutes to 4 (right) of the 5-minute testing period. The abscissae represent the lag from an initial B→S conjunction and the beginning of the 3-second window of time. The ordinates represent the average deviation from chance of the probability of occurrence of a body movement in a 3-second window of time. The two dotted lines of each graph represent plus and minus twice the standard error (mean of the 14 SEM calculated for each session).

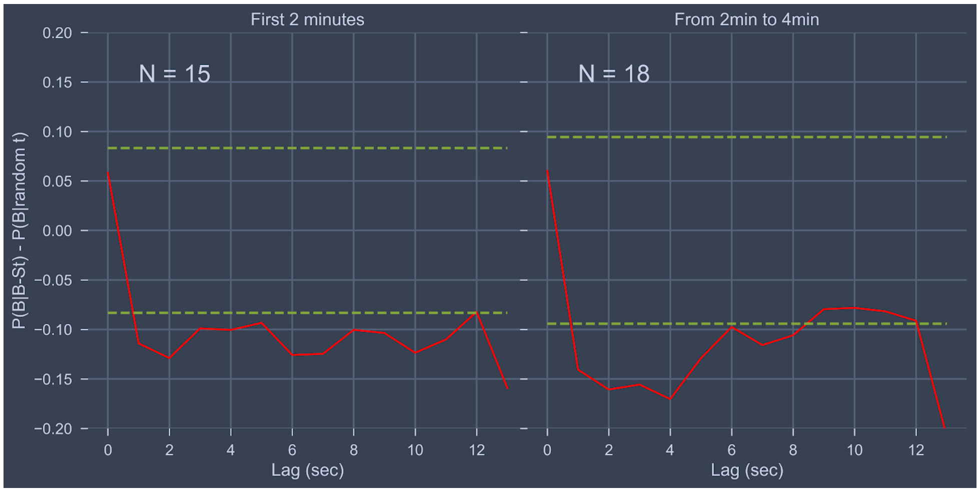

Figure 2 - Means and standard error of the means of the deviation from chance (calculated from the master bracelet) of P(B|B-St) during the first 2 minutes (left) and from 2 minutes to 4 minutes (right) of the period of test in the contingent group. The abscissae represent the lag from an initial B→S conjunction and the beginning of the 3-second window of time.

I expected to see similar patterns to the ones observed by Watson et al. in their experiment: an inhibition (after about 1-2 seconds) and an excitation (from around 6 to 8 seconds). Surprisingly, the curves are almost always negative, meaning that after an occurrence of a B→S conjunction there is a systematic inhibition of master bracelet movements for at least 16 seconds (13 seconds plus the 3 seconds window size). However, it could be that this finding is an artefact, due to the fact that P(B|random t) is calculated from the 5-minute testing period and not from the baseline as Watson did. As a result, P(B|random t) is not the probability of a body movement in any context but the probability of a body movement in a specific context in which a reward is associated with these body movements. A way to avoid this bias is to calculate P(B|random t) from the observed rate of body movement of the slave bracelet during the 5-minute period of test. Indeed, the slave is as salient as the master. However, the motions of the slave arm are not directly reinforced, contrary to those of the master arm.

The results with P(B|random t) calculated from the observed rate of body movement of the slave bracelet are plotted in the two figures below:

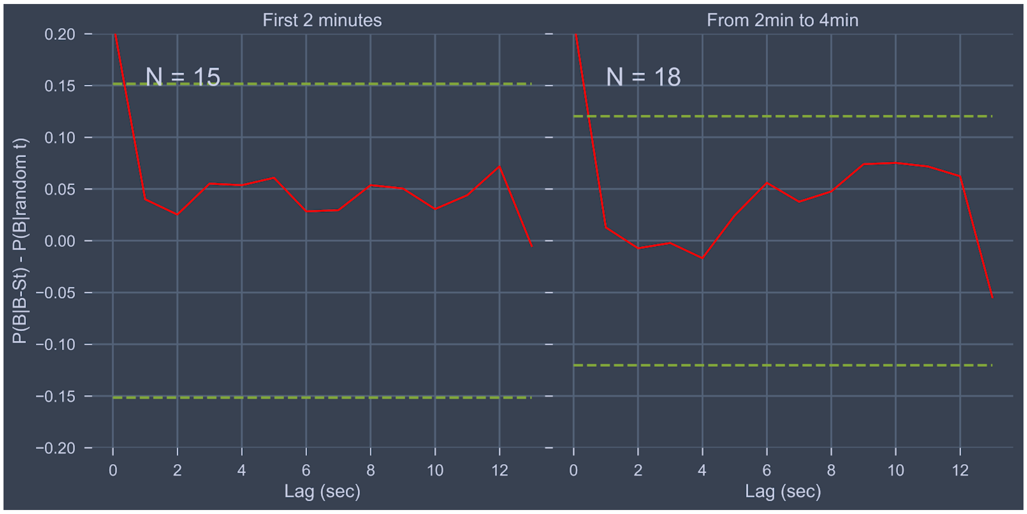

Figure 3 - Means and standard error of the means of the deviation from chance (calculated from the slave bracelet) of P(B|B-St) during the first 2 minutes (left) and from 2 minutes to 4 minutes (right) of the period of test in the contingent group. The abscissae represent the lag from an initial B→S conjunction and the beginning of the 3-second window of time.

In this case, we see that the probability of a body movement given the prior occurrence of a B→S conjunction does not differ from chance for both periods. Notice that the four curves (figures 2, 3) are positive at the lag 0 second. The effect is consistent whatever the condition, and lasts less than one second as the peaks disappear at the lag 1 second. Given these features, the peaks are very probably due to the inertia of the babies' motions.

Discussion

First, in order to investigate the tools designed by Watson we reproduced them as closely as possible. Contrary to our expectations, we had no evidence that infants produced any distinctive behavioral pattern when exposed to our contingency as Watson found (see figures 1E, 1F).

The conclusion would have been utterly different if the error bars were nearer to the zero axis. To make sure that I was not making a mistake, I tried an alternative way of estimating them. I computed the 95% confidence intervals for each graph using bootstrapping with replacement (see Appendix D for details). These new error bars were coincident with the old, corroborating our conclusion.

The most questionable point about the analysis we have done is the value of P(B|random t). Until now, I have assumed that P(B|random t) has a Poisson distribution as Watson did. Watson defined P(B|random t) as the probability of occurrence of a body movement at random in a neutral context. P(B|random t) was a measure of the baby's restlessness. As we had no baseline period, I defined P(B|random t) as the probability of finding at least one occurence of B in a t second bin chosen at random in the experiment. Both definitions are informative but they are different. In our case, it does not make sense to use a Poisson distribution as the events B do not occur: (1) with a constant rate over the period of test for all the babies (see Appendix E for a quick overview of the diversity of patterns existing) and (2) independently of the event before. Since the conditions to use Poisson's law were not valid, I decided to go back to the definition of P(B|random t) chosen above. I calculated P(B|random t) using sampling with replacement throughout the 5-minute period of test (50000 samples of span 3 seconds were drawn for each subject) to obtain a more realistic appraisal of its value. The new values of P(B|random t) were lower than the old ones and the curves were above the confidence intervals. This last result leads to confusion. The three attempts to estimate P(B|random t) gave rise to three different results. We can not assess the success of the contingent group without a better understanding of this variability.

To interpret the shapes of the mean curves in figure 2, I looked at the individual curves producing them (see Appendix F). It is remarkable that they differ considerably from one another. Consequently, the mean curves are not representative of any real behavioral strategy. As Watson did not show the individual curves in his paper we can not reject the possibility that his data was equally diverse. We should keep this in mind when comparing his predictions to our results.

Because we found a great variability among the individual curves it will be interesting to look at the results for individual babies, to see if, within an individual baby's behaviour we are able to find evidence that it has understood the contingency. This analysis at the individual level may enable us to quantify more precisely the success of the babies since if the majority of them do not detect the contingency we will be able to know how many did.

For the analyses to be presented below, it will also be particularly relevant for us to keep working on the activity of the babies in very short periods of time. Measures in short windows of time are less sensitive to parasitic behaviors such as non-nutritive sucking, playing with the feet, looking toward the parent, visual exploration of the room etc., than measures all along the experiment. Watson's probabilistic reasoning was an example of how to select those short windows of time along the experiment. Continuing to use his methodology at the individual level can bring to light strategies that would otherwise become mixed up with parasitic behaviors. For example, consider what would happen if a baby very rapidly understands the role of the master after only a few trials, and then goes on to test the slave bracelet by moving it a great number of times before subsequently abandoning it and becoming interested in its mother's hands. A measure comparing this whole period with the end of the experiment, in which let's say the baby sucks the bracelets, will show neither an increase of the motions of the master nor an increase of the motions of the slave. On the contrary, a measure able to focus on the specific period in which the baby takes an interest in the slave bracelet can show the contrast existing between this period and the rest of the experiment.

The replication of Watson's work did not solve our research question. However, his approach represents a new theoretical framework waiting to be further explored. The following analysis is to a great extent inspired from Watson's. My purpose was to design a simple tool to distinguish babies who detected the contingency from those who did not.

5. Second analysis

In his work, Watson showed a reaction to a B→S conjunction different from chance and different from both reactions to B and S separately *at the group level*. It led me to look at the same question *at the individual level* (see discussion of the first analysis above). For a particular baby from the contingent group, is the probability of body movements following a B→S conjunction (P(B|B-St)) different from the probability of body movements at random (P(B|random t))? If so, can we observe an analogous difference within the control group?

Adaptation of Watson’s analysis to our experiment

To answer this question, I created a binary indicator. In the contingent group, instead of looking at P(B|B-St) at different lags following a prior B→S conjunction, I chose to measure Bs one second after a prior B→S conjunction only. I did not measure Bs immediately after a prior B→S conjunction to avoid a bias due to the inertia of the baby’s motions (see points with lag = 0 seconds in figures 2, 3) and I did not extend the lag more so as to have the most direct measure of the body movements occurring after the reference B→S conjunction.

As a result, I chose P(B|B-St) to be the probability of at least one occurrence of B in a time range t (t = 3 seconds) starting 1 second after a B→S conjunction. To calculate P(B|B-St) I divided the number of windows of time t in which at least one occurrence of B happened following a B→S conjunction with a 1-second lag, by the number of windows of time t in which it could have happened, that is the number of windows of time t following a B→S conjunction. As I did in the first analysis, I applied an arbitrary criterion of at least 5 target event (B) and 5 base events (B→S) to enter into the analysis in order to be able to calculate P(B|B-St).

In the contingent group, in order to take into account the broadest possible range of strategies and to know more about arm differentiation (recall that one of our research question was “Will babies realize that the contingency is determined by only one of their arm?”) I computed 3 extra indicators analogous to P(B|B-St): P(B’|B-St), P(B|B’t) and P(B’|B’t) with B’ being “a body movement of the slave bracelet where the acceleration is greater than the threshold”. These 3 indicators were computed using the same methodology that I described above to compute P(B|B-St). I applied an arbitrary criterion of at least 5 target events (B or B’) and 5 base events (B→S or B’) to enter into the analysis in order to be able to calculate P(B’|B-St), P(B|B’t) and P(B’|B’t). Notice that a B’→S conjunction did not exist as the slave bracelet did not trigger the contingency.

In the control group, I computed the following indicators (analogous to the ones Watson used): P(B|Bt), P(B’|B’t), P(B|St) and P(B’|St), the probabilities of at least one occurence of B or B’ in a time range t (t = 3 seconds) starting 1 second after a prior occurrence of B, B’ or S. These 4 indicators were computed using the same methodology that I described above to compute P(B|B-St). I applied an arbitrary criterion of at least 5 target events (B or B') and 5 base events (B, B' or S) to enter into the analysis in order to be able to calculate P(B|Bt), P(B'|B't), P(B|St) and P(B'|St). Notice that in the control group the functions of the master and the slave bracelets are equivalent. I kept the notations B and B' to differentiate between arms only.

In both contingent and control groups, P(B|random t) and P(B'|random t) were the chance expectancies of observing at least one occurrence of B or B' in a time span t (t = 3 seconds) following a randomly specified point in time during the 5-minutes of experiment. P(B|random t) and P(B'|random t) were computed using sampling with replacement throughout the 5-minute period of test (50000 samples of span 3 seconds were drawn for each subject). I also computed the 95% confidence intervals of P(B|random t) and P(B'|random t) from this sampling (see Appendix D for details).

In order to have the best chance of capturing the moment during the 5-minute period of test at which the baby realized the presence of a contingency, I decided first to locate during this whole period that particular 2-minute period in which the baby had the highest number of Bs. Then I computed the indicators for both the contingent and the control groups for each baby for this period. The duration of 2 minutes was a compromise between detecting temporary reactions to the contingency and allowing for minimum numbers of B and B'. After that, since I wanted to take into account every possible strategy including the exploration of the contingency with the slave bracelet, I also computed my indicators in the 2-minute period in which the baby had the highest number of B's (see Appendix G). Finally, as there were babies without enough occurrences of B and B' in these two 2-minute periods, I looked at the whole 5-minute period (see Appendix H).

For each of my indicators, I computed the 95% confidence intervals using bootstrapping with replacement (see Appendix D for details) to compare them with chance.

Results

The following figure contains 18 graphs representing the 18 babies of the contingent group. I titled each graph with the alias of the corresponding baby (I did the same for figure 5). The graphs cover, for each baby, the first 2 minutes with the highest number of occurrences of B within the 5-minute testing period. Each graph is divided into two blocks representing (1) the probability of occurrence of B (given the conditions mentioned above) on the left (warm block) ; and (2) the probability of occurrence of B’ (given the conditions mentioned above) on the right (cold block). The left block is made up of P(B|B-St) (red), P(B|B’t) (tomato) and P(B|random t) (yellow). The right block is made up of P(B’|B-St) (dark blue), P(B’|B’t) (light blue) and P(B’|random t) (green). The error bars (black) represent the 95% confidence intervals of the means of the samples. At the top of the graphs are annotated the sizes of the samples, that is the number of body movements, B and B’, within the 2-minute period of the master (M) and the slave (S) bracelets respectively. The tomato bars and cold blocks are missing in graphs titled “amebru”, “eliger” and “solcha” because a minimum of 5 occurrences of B’ was required to compute P(B|B’t), P(B’|B-St) and P(B’|B’t).

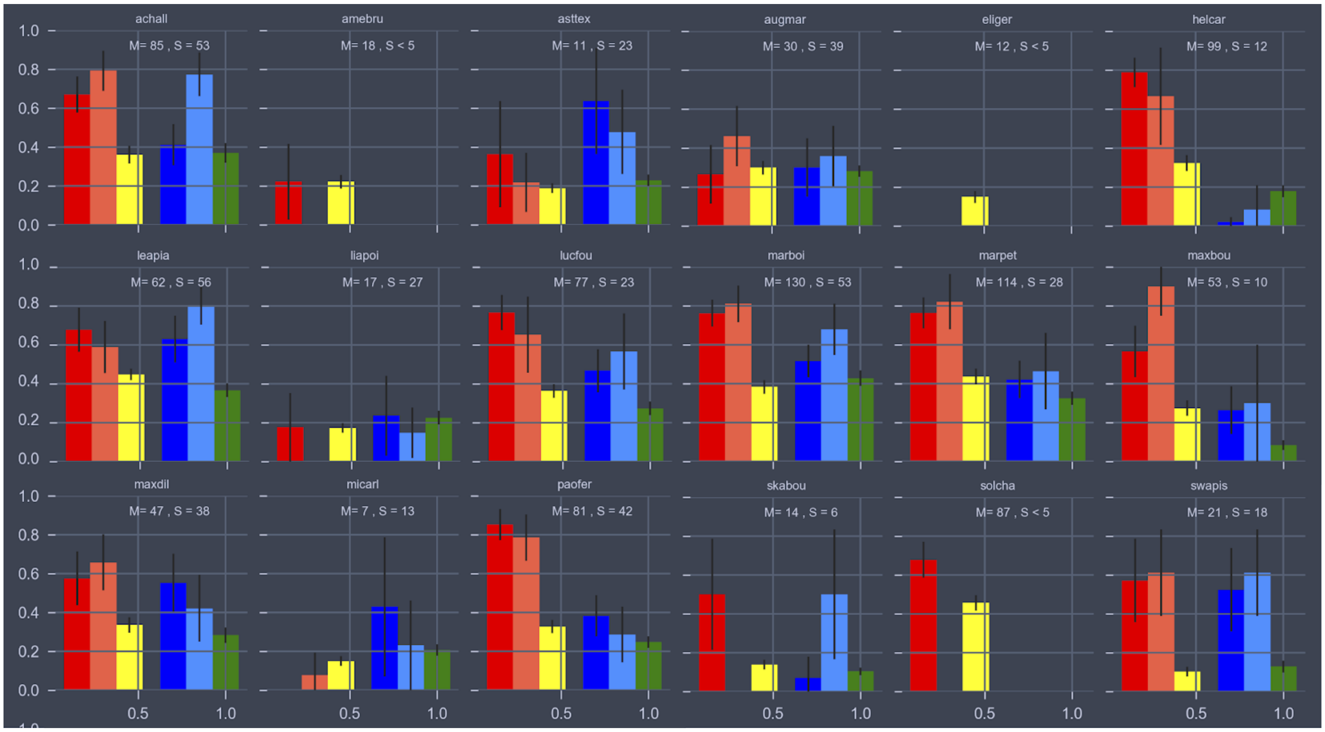

Figure 4 - P(B|B-St) (red), P(B|B’t) (tomato), P(B|random t) (yellow), P(B’|B-St) (dark blue), P(B’|B’t) (light blue) and P(B’|random t) (green) for each baby in the contingent group during the first 2 minutes with the highest number of occurrences of B within the period of test. The error bars (black) are the 95% confidence intervals for each value. M and S are the number of occurrences of B and B’ respectively from which P(B|B-St), P(B|B’t), P(B’|B-St) and P(B’|B’t) are computed.

Figure 5 contains 11 graphs representing the 11 babies of the control group. As in Figure 4, the graphs cover, for each baby, the first 2 minutes with the highest number of occurrences of

B within the 5-minute testing period. Each graph is divided into two blocks representing the probability of occurrence of body movements for each arm. Notice that the blocks can be switched indifferently since, in the control group, there were no functional differences between the bracelet worn on one arm versus the bracelet worn on the other arm. The left block is made up of P(B|Bt) (caramel), P(B|St) (blood orange) and P(B|random t) (yellow). The right block is made up of P(B'|B't)(violet), P(B'|St) (flashy violet) and P(B'|random t) (green). Similarly to Figure 4, the error bars (black) represent the 95% confidence intervals of the means of the samples. At the top of the graphs are annotated the number of occurrences of the responsive stimulus (N) and the sizes of the samples (see above) within the 2-minute period. In graphs titled "elorib", "jeafau" and "soalam" bars are missing because a minimum of 5 occurrences of B or B' was required to compute P(B|Bt), P(B|St), P(B'|B't) or P(B'|St).

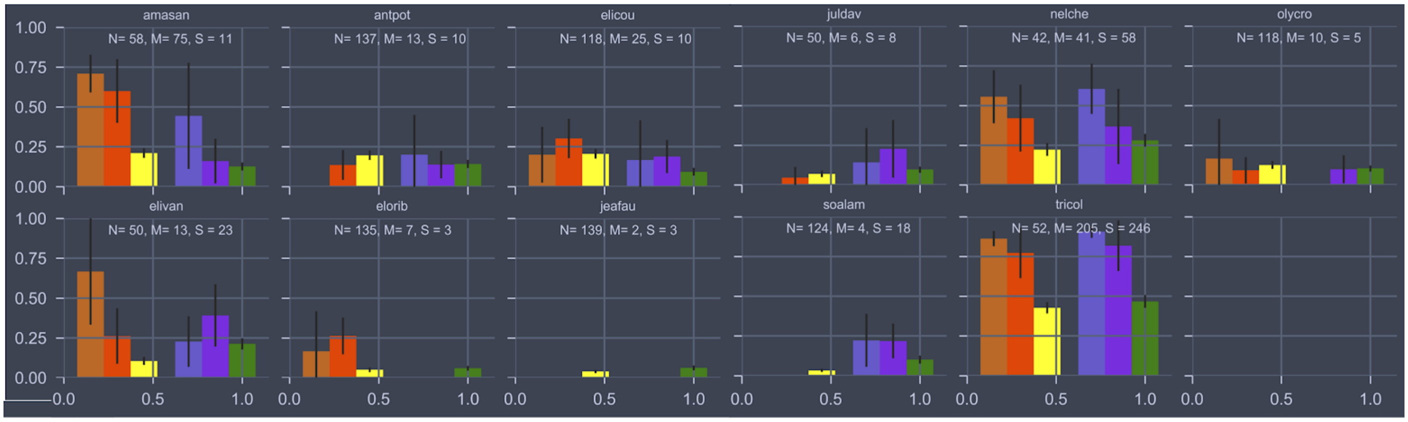

Figure 5 - P(B|Bt) (caramel), P(B|St) (blood orange), P(B|random t) (yellow), P(B'|B't) (violet), P(B'|St) (flashy violet) and P(B'|random t) (green) for each baby in the control group during the first 2 minutes with the highest number of occurrences of B within the period of test. The error bars (black) are the 95% confidence intervals for each value. M and S are the numbers of occurrences of B and B' respectively from which P(B|Bt), P(B|St), P(B'|B't) and P(B'|St) are computed.

In the contingent group, I expected to find babies showing statistical differences between our indicators and chance (the yellow and green bars), that is no overlap between their 95% confidence intervals. On the contrary, in the control group, I expected to find no statistical difference between our indicators and chance (yellow and green bars) for every baby with no exceptions. Otherwise, particularly if babies from the control group reacted to the responsive stimulus, I expected to find greater probabilities of body movements following S in the subgroup with the greatest number of S.

Results of the contingent group

There are three babies (amebru, augmar, liapoi) with neither statistical difference between P(B|B-St) or P(B|B’t) and (P(B|random t) nor statistical difference between P(B’|B-St) or P(B’|B’t) and P(B’|random t). We can reasonably confirm that *within the chosen 2-minute period* the contingency does not have any impact on B and B’. Either those babies were not interested in the stimulus or they were interested in the stimulus but did not make the connection with their body movements. In both cases, they did not detect the contingency *within the chosen 2-minute period*. Interestingly, babies amebru and augmar show the same results in the 2-minute period with highest number of B’s (see Appendix G) and in the whole 5-minute period of test (see Appendix H) confirming our first conclusion, but liapoi does not and we will reconsider it later.

There are twelve babies (achall, helcar, leapia, lucfou, marboi, marpet, maxbou, maxdil, paofer, skabou, solcha, swapis) with P(B|B-St) statistically greater than chance, namely P(B|random t). We can reasonably infer for those babies that B or S or B→S have an impact on B. In order to clear up the role of each of these three events we need to look at the results of the control group. The most indisputable result would be the following: in the control group, P(B|Bt) and P(B|St) are equal to P(B|random t) meaning that B is neither reinforced nor inhibited by prior occurrences of B or S respectively. Then, for the twelve previous babies from the contingent group we would be able to confirm that B→S has an impact on B: we assume that the babies perceived the contingency *at least within the chosen 2-minute period*. For seven of those twelve babies (achall, helcar, lucfou, marboi, marpet, maxbou, paofer) P(B|B-St) was statistically greater than P(B’|B-St). If this result turns out to be specific to the contingent group and is no found in the control group we assume that those seven babies distinguished the master bracelet from the slave bracelet since they made use of them differently. For babies with no difference between P(B|B-St) and P(B’|B-St) our indicators do not allow to conclude for arm differentiation: either babies moved both arms together after every B→S conjunction or they moved the master arm following half of the B→S conjunctions and the slave arm for the other half.

There is one baby (asttex) with ambiguous results: P(B’|B-St) statistically greater than P(B’|random t) although P(B|B-St) smaller than P(B|random t). This baby may be wondering if the slave bracelet can trigger S. This hypothesis is strengthened by the following result: P(B'|B-St) and P(B'|B't) greater than chance in the 2-minute period with greatest number of B's (see Appendix G). Then, if the results from the control group show no excitatory reaction following a body movement (P(B|Bt) > P(B|random t) or P(B'|B't) > P(B|random t)) we assume asttex detected the contingency. As, in figure 9, P(B'|B't) is also greater than chance for liapoi we assume it also detected the contingency.

Finally, there are two babies (eliger, micarl) showing an inhibitory reaction to the contingency (P(B|B-St) < P(B|random t)). To interpret this result, we can look at the 2-minute periods in with highest number of B's (see Appendix G) and at the whole 5-minute period (see Appendix H). As there is no statistical difference between chance and our indicators for both eliger and micarl, we confirm that neither eliger, nor micarl detected our contingency.

Results of the control group

Recall that in the control group, babies were exposed to S at two different rates (see Methods). I will refer to the subgroup in which babies were exposed to S for a tenth of the 5-minute testing period as the *low activity control group* to contrast with the *high activity control group* in which babies were exposed to S for a third of the 5-minute testing period. Babies amasan, elivan, juldav, nelche and tricol belong to the low activity control group and babies antpot, elicou, elorib, jeafau, olycro and soalam belong to the high activity control group. Notice that, within a subgroup babies were not exposed to strictly the same amount of S (N varies depending on the baby). This is due to our method to generate random occurrences of S (see Appendix B for more about the responsive stimulus).

In the high activity control group (6 babies), we found the results we expected, namely no statistical difference between our indicators and chance for every baby, with no exceptions. Besides, notice that the number of body movements is extremely low. In the low activity control group (5 babies), there is one baby showing the expected results (juldav) and the others show patterns similar to the ones of the contingent group. These results suggest an effect of the number of occurrences of the responsive stimulus (N) to be confirmed with more babies.

Discussion

This second analysis was done to investigate the individual reaction of 6-month old babies to our contingency. In order to expose babies from both contingent and control groups to the same amount of arousal, we divided the control group into two subgroups. The behaviors of the babies in the low activity control group were remarkably different from those of the high activity control group. Comparing the contingent group to the high activity control group suggests that babies of age 6 months are able to detect our contingency whereas, comparing the contingent group to the low activity control group suggest that we can not conclude anything from our indicators since babies from both groups show similar behavioral patterns. Given these opposing results, should we conclude that, in the end, we are in a deadlock?

My answer is no, we are not. My hypothesis is that, in the low activity control group, babies move their arms to test a possible contingency between their movements and the responsive stimulus. Then, the similarities between the bar charts of the contingent group and the low activity control group do not mean that our indicators are wrong but rather that in both conditions babies suspected the existence of a contingency between their motions and the stimuli. Recall that the responsive stimulus in fact had six variants (see Appendix B). It is likely that babies in the low activity control group were moving when the responsive stimuli occurred. After 30 seconds for example, a typical baby had done several movements and all the six variants of responsive stimuli had occurred several times. If the baby remembered that the responsive stimuli had previously occurred, it could try to reproduce its previous actions. This would happen even if the baby is part of the low activity control group, leading it to explore a possible contingency. If this is true, why did babies in the high activity control group not behave similarly? Here, the analysis of the video recordings is helpful. Before watching the recordings, I anonymized the files. After that, I used an observation grid to categorize the babies' behaviors. Whereas I could be misled by the babies from the low activity control group (I needed a few seconds to make sure that they did not belong to the contingent group), the responsive stimulus in the high activity control group occurred so frequently that it took me only a few seconds to ascertain that the babies were not causing it. Thus we can imagine that babies in the high activity control group also perceived that the responsive stimuli occurred without causal link to any of their behaviors. According to this hypothesis, babies of age 6 months are able to distinguish between a situation in which they can be the cause of the responsive stimulus and a situation in which they can not. This is a prerequisite to contingency detection and a positive indicator that our experimental design leads babies to explore our contingency.

Our goal was to answer two questions: (1) Are 6-month-old infants able to detect our contingency? (2) Will they realize that this contingency is determined by only one of their arms? At first sight, the second analysis, in particular the comparison between the contingent and the low activity control groups, showed that babies can behave similarly when they have no control at all or a total control of the responsive stimulus. However, if we look closely at the graphs, it appears necessary to distinguish the comparisons intrablock (P(B|B-St) vs P(B|random t) or P(B'|B-St) vs P(B'|random t) for the contingent group, and P(B|Bt) or P(B|St) vs P(B|random t) for the control group) from the comparisons interblock (P(B|B-St) vs P(B'|B-St) for the contingent group, and P(B|Bt) vs P(B'|B't) or P(B|St) vs P(B'|St) for the control group). Indeed, the differences intrablock in the contingent group are similar to the ones in the low activity control group but this is not true for the differences interblock. In the low activity control group, one baby (amasan) shows a statistically significant difference between P(B|St) and P(B'|St) whereas in the contingent group, seven babies show a statistically significant difference between P(B|B-St) and P(B'|B-St). Two hypotheses are possible: either amasan is an outsider and the coming experiments will confirm the ability to differentiate between their arms for 6-month old babies in our experiment, or others babies in the low activity control group will show "arm differentiation" and my indicators are meaningless. In fact, it is likely that we observe a similar rate of babies showing arm differentiation in both contingent and low activity control group. In our experiment babies were rewarded equally for movements of the master arm only and for movements of the whole body. Many studies show limb differentiation in infants as early as 3-4 months of age even though babies were equally rewarded for movements of the connected limb or movements of the whole body (Rovee-Collier et al., 1978; M. Angulo-Kinzler, 2001; Heathcock et al., 2005; Watanabe & Taga, 2006, 2009), but in these studies the babies' arms or leg was linked to the mobile with a ribbon. The absence of local tactile stimulus in our experiment, providing additional information about the contingency, could explain that we observe a similar rate of babies showing no arm differentiation in both contingent and low activity control group. Furthermore, it may be likely that visual, auditory and tactile feedback are not equally informative or that their combination provides supplementary knowledge. Maybe the main limitation to this second analysis is the duration of the experiment. In his control group, Watson showed a change in babies' behaviors between the first 2-minute period and the second 2-minute period of his experiment. At the beginning of the experiment, babies reacted to the stimulus, while after 2 minutes, they stop moving according to the occurrences of the stimulus. We might obtain a similar result if babies had more time to explore the apparatus in our control group. This might help to to clarify our results.

6. Conclusion

The present work was done to develop probabilistic tools inspired from the literature to gauge if 6-month old babies can detect an artificial contingency between one of their arms and an auditory and visual responsive stimulus. The tool that I adapted from Watson's work to gauge babies' behavior at the individual level takes into account a wide range of strategies and, as a result, differentiates babies that explore the experimental set-up from the ones that do not pay attention to it or (in the contingent group) do not understand how it works. This tool is finer than the usual indicators used to assess babies' success, leading to more ambiguous results as it reveals that (contrary to what would be easier to interpret) babies explore their environment with similar behaviors in similar circumstances.

7. References

Angulo-Barroso, R. M., Peciña, S., Lin, X., Li, M., Sturza, J., Shao, J., & Lozoff, B. (2017). Implicit learning and emotional responses in nine-month-old infants. *Cognition and Emotion*, *31*(5), 1031–1040.

Bloom, J. & Orloff, J. (2014). *Bootstrap confidence intervals*. (n.d.). Retrieved from: https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading24.pdf

Butko, N. J., & Movellan, J. R. (2010). Detecting contingencies: An infomax approach. *Neural Networks*, *23*(8–9), 973–984.

Corbetta, D. (2018). *Reach-to-Grasp Behavior: Brain, Behavior, and Modelling Across the Life Span*.

Fagen, J. W., Morrongiello, B. A., Rovee-Collier, C., & Gekoski, M. J. (1984). Expectancies and Memory Retrieval in Three-Month-Old Infants. *Child Development*, *55*(3), 936–943. Retrieved from JSTOR.

Heathcock, J. C., Bhat, A. N., Lobo, M. A., & Galloway, J. (Cole). (2005). The Relative Kicking Frequency of Infants Born Full-term and Preterm During Learning and Short-term and Long-term Memory Periods of the Mobile Paradigm. *Physical Therapy*, *85*(1), 8–18.

Jacquey, L., Fagard, J., Esseily, E., & O'Regan, K. J. (in preparation). *Detection of sensorimotor contingencies in infants before the age of one year: a comprehensive review*.

Jacquey, L., Popescu, S., Vergne, J., Fagard, J., Esseily, E., & O'Regan, K. J. (submitted). *Development of body knowledge as measured by arm differentiation in infants: from global to local?*

M. Angulo-Kinzler, R., & L. Horn, C. (2001). Selection and memory of a lower limb motor-perceptual task in 3-month-old infants. *Infant Behavior and Development*, *24*, 239–257.

Movellan, J. R. (2005). An Infomax Controller for Real Time Detection of Social Contingency. *Proceedings. The 4nd International Conference on Development and Learning, 2005.*, 19–24. Osaka, Japan: IEEE.

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, *11*(2), 265–286.

Richters, D. P., & Eskew, R. T. (2009). Quantifying the effect of natural and arbitrary sensorimotor contingencies on chromatic judgments. *Journal of Vision*, *9*(4), 27–27.

Rovee, C. K., & Rovee, D. T. (1969). Conjugate reinforcement of infant exploratory behavior. *Journal of Experimental Child Psychology*, *8*(1), 33–39.

Rovee-Collier, C. K., Morrongiello, B. A., Aron, M., & Kupersmidt, J. (1978). Topographical response differentiation and reversal in 3-month-old infants. *Infant Behavior and Development*, *1*, 323–333.

Scharoun, S. M., & Bryden, P. J. (2014). Hand preference, performance abilities, and hand selection in children. *Frontiers in Psychology*, *5*.

Tiernan, C. W., & Angulo-Barroso, R. M. (2008). Constrained Motor-Perceptual Task in Infancy: Effects of Sensory Modality. *Journal of Motor Behavior*, *40*(2), 133–142.

Watanabe, H., & Taga, G. (2006). General to specific development of movement patterns and memory for contingency between actions and events in young infants. *Infant Behavior and Development*, *29*(3), 402–422.

Watanabe, H., & Taga, G. (2009). Flexibility in infant actions during arm- and leg-based learning in a mobile paradigm. *Infant Behavior and Development*, *32*(1), 79–90.

Watson, J. S. (1979). Perception of contingency as a determinant of social responsiveness. In *The Origin of the infant's Social Responsiveness* (Lawrence Erlbaum Associates, pp. 33–64). Thoman, Evelyn B.

Watson, J. S. (1984). Memory in learning: Analysis of three momentary reactions of infants. In *Comparative Perspectives on the Development of Memory* (Lawrence Erlbaum Associates, pp. 159–179). Available at:

https://www.dropbox.com/s/iw3h99ywk2q1rqs/Watson%20contingency%20pap37.pdf?dl=0

Watson, J. S. (1985). Contingency perception in early social development. In *Social Perception in Infants* (Ablex Publishing Corporation, pp. 157–176). Norwood: Fox, Nathan A.; Field, Tiffany.

Appendix A. Detailed presentation of the material

The experimental set-up consisted of 3 microcontrollers and a computer gathering information. The writer microcontroller (W) was connected to the computer through a USB wire (black solid line). The master (M) and the slave (S) microcontrollers were placed in transparent boxes attached to the wrists of the subject with velcro ribbons (the bracelets). W indicated to M and S the beginning and the end of the test session (green dotted lines). The microcontrollers sent and received data through wifi connections (dotted lines). M triggered its activations (sound and light) but also those of S (red dotted line). Both microcontrollers (M and S) sent their data to W all along the session (blue dotted lines).

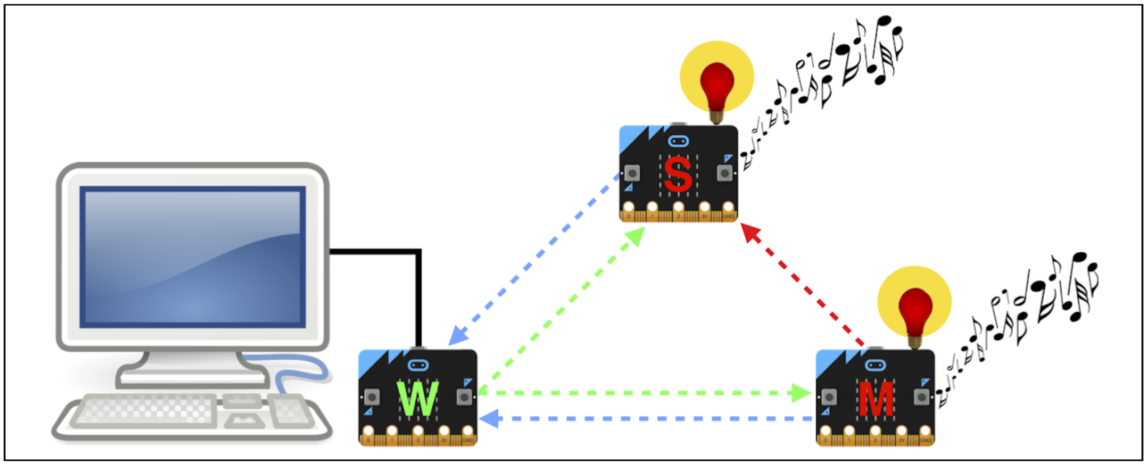

Figure 6 - The experimental set-up. 3 microcontrollers communicated through wifi connections (dotted lines). One of them (W) sent information to the computer through a USB wire (solid line).

M and S were programmed to send their data every 100 milliseconds (expected sending frequency : 10 Hz). They sent their time-stamp, their threshold (above which activation occurred), the accelerations along the 3 axes of the on-board accelerometers and whether or not they were activated.

Appendix B. The responsive stimulus

A responsive stimulus lasted 500 milliseconds. It was made up of a melody of five notes (the sound) and a succession of five visual patterns (the light). There were six different responsive stimuli, namely six different melodies (made up of 5 notes chosen in a library containing 3 notes plus silence) combined with six different successions of visual patterns (made up of 5 patterns chosen in a library containing 3 patterns plus lights off). For each occurrence of the responsive stimulus, the latter was drawn at random among the six possible responsive stimuli.

In the control group, to generate the occurrences of the responsive stimulus at random, we created a list of twenty binary items (True/False). The value of each item was "True" with the probability of 0.11 or 0.29 according to the subgroup of the baby (see methods). Then, during the 5-minute period of experiment, every 500 milliseconds, the algorithm drew at random from the list of binary items. If "True" was drawn, a responsive stimulus occurred. We had to create a list of binary items in advance rather than drawing every 500 milliseconds because of the limited memory capacity of the microcontrollers. The duration 500 milliseconds was chosen because it is the duration of the responsive stimulus. Consequently, in the contingent group, if a baby triggered the contingency continuously, the responsive stimulus would have occurred every 500 ms maximum.

Appendix C. Inclusion criteria

In the contingent group, 3 babies were excluded because of technical problems during the experiment (no data recorded for 2 babies and dead battery for one baby).

For both groups, there were 4 additional inclusion criteria. They were chosen to exclude the babies with low quality data and thereby avoid comparisons between babies with nonequivalent data quantity (all along the test or locally).

First, we need to check for the regularity of the sendings. We are interested in the patterns of stimuli babies may produce. Each stimulus lasts 500 ms. So the highest frequency for a pattern is $\frac{1}{0.5s} = 2\,Hz$. The Nyquist–Shannon sampling theorem gives the minimum sending frequency to guarantee the integrity of the signal reconstruction : $2 * 2\,Hz = 4\,Hz$. The average sending frequencies observed for the master and the slave microcontrollers are 8.94 Hz >> 4 Hz and 9.05 Hz >> 4 Hz. The Nyquist–Shannon sampling theorem is observed. Nevertheless, to compare our subjects we need them to have an equivalent amount of data. We set the minimum sampling frequency at 9 Hz for both microcontrollers and exclude subjects with less than 90% of their samples sent at least at this frequency (2 babies in the contingent group).

Then, we check for the homogeneity of the sample distribution all along the test that is, whether there are periods with no sendings. We set at 2.5 s the maximum period with no sendings tolerated for both microcontrollers. No subject is excluded according to this criterion.

Finally, we exclude babies who did not complete the 5-minute period of test (7 babies in the contingent group and 1 baby in the control group) and the ones who did not trigger the responsive stimulus at least 5 times (1 baby in the contingent group and on baby in the control group). This criterion was established to calculate P(B|B-St).

Appendix D. Bootstrapping to compute confidence intervals

To compute the confidence intervals (CI) using bootstrapping in my analyses I applied a methodology described by Jeremy Orloff & Jonathan Bloom in their MIT online class “Introduction to Probability and Statistics” (*Bootstrap confidence intervals*, n.d.).

The methodology is based on the following principle: Suppose I have a sample of n data points $(x_1, x_2, ..., x_n)$ from a population and u is a statistic computed from this sample. An empirical bootstrap sample is a resample of the same size n $(x_1^*, x_2^*, ..., x_n^*)$ and $u*$ the statistic computed from the resample. The bootstrap principle says that the variation of u is well-approximated by the variation of u^*.

The methodology to compute the 95% CI of the mean $\bar{x}$ of a sample of size n $(x_1, x_2, ..., x_n)$ is:

(1) To draw n times in $(x_1, x_2, ..., x_n)$ with replacement (after each draw x_i is put back in the sample) to obtain a resample of size n $(x_1^*, x_2^*, ..., x_n^*)$.

(2) To compute the mean $\overline{x^*}$ of the resample.

(3) To repeat (1) and (2) N times to obtain a list of the means of the N resamples L = [$\overline{x_1^*}$, …, $\overline{x_N^*}$].

(4) To sort L in ascending order.

(5) To select the 2.5th percentile and the 97.5th percentile of the list. If N = 1000, it means to choose the 25th and 975th elements of the list. Else, it means to chose the largest integer value less than or equal to $\frac{2.5*N}{100}$ and the smallest integer value greater than or equal to $\frac{97.5*N}{100}$.

(6) To subtract the 2.5th percentile to the 97.5th percentile: this is the size of the CI.

Appendix E. Rates of activation per minute over time

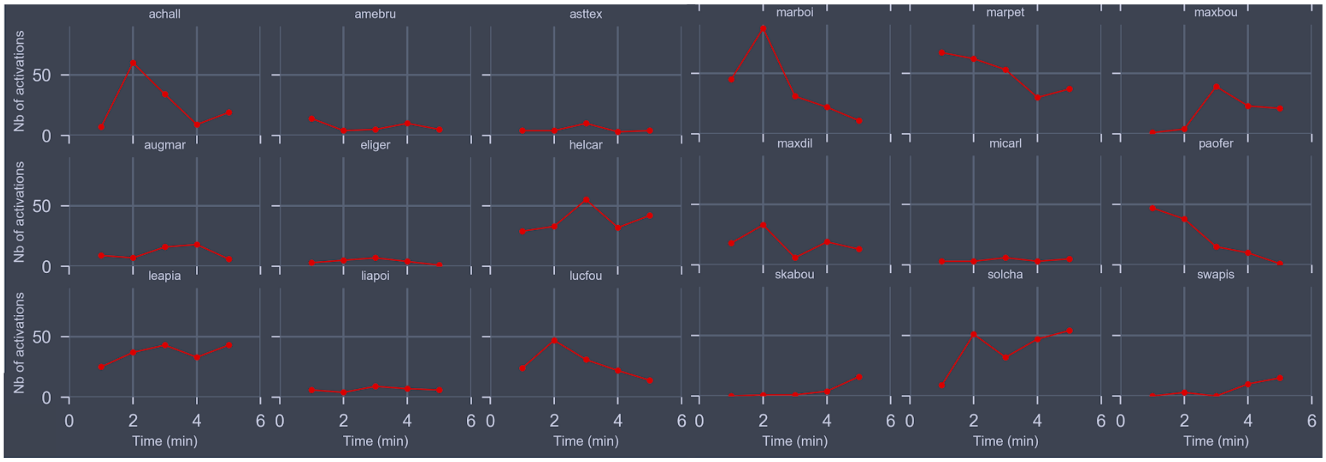

Figure 7 - Number of body movements B per minute over the 5 minutes of experiment for each subject in the contingent group. The majority of the graphs are far from a constant rate.

Appendix F. Individual patterns producing Figure 2

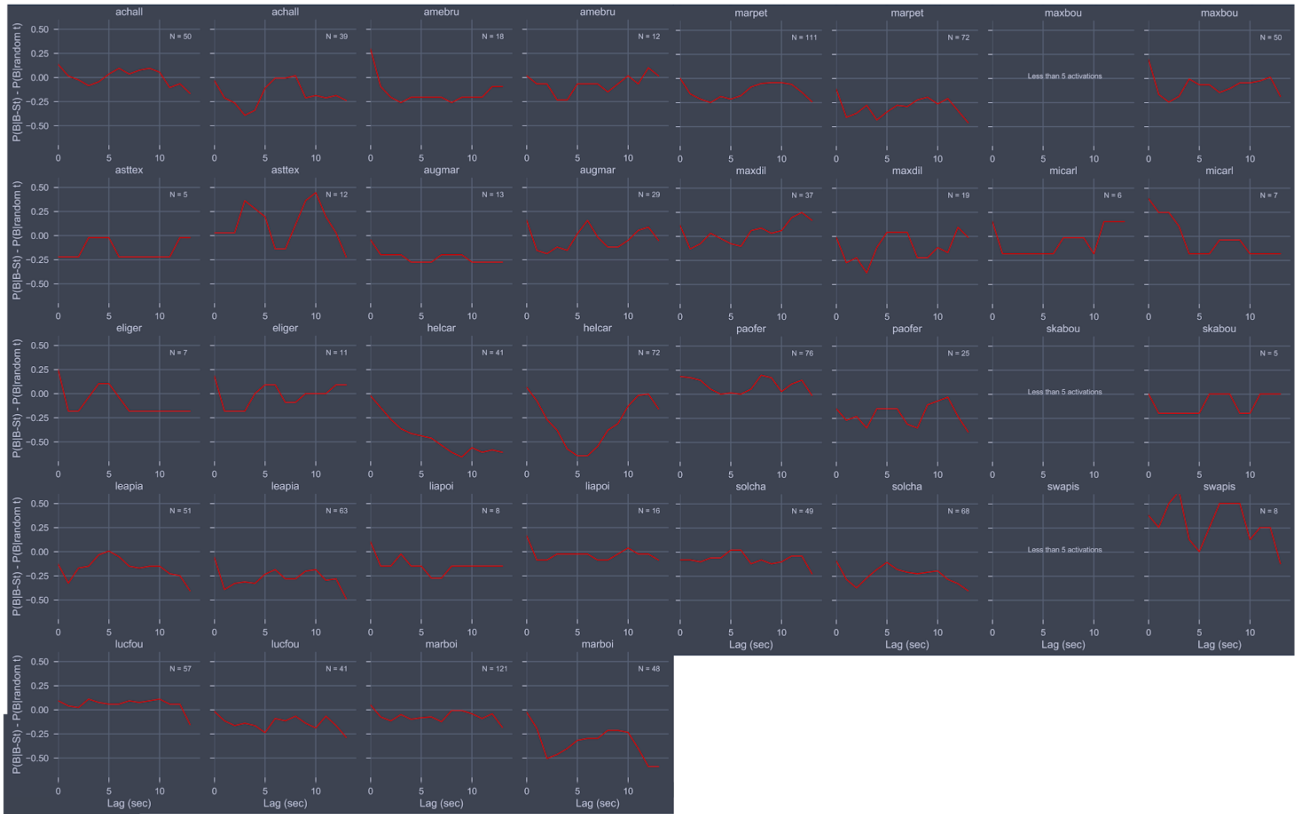

Figure 8 - P(B|B-St)-P(B|random) for each baby during the first 2 minutes (left) and from 2 minutes to 4 minutes (right) of the period of test in the contingent group. The abscissae represent the lag from an initial B→S conjunction and the beginning of the 3-second window of time.

Appendix G. Second analysis in the first 2 minutes with the greatest number of body movements of the *slave* bracelet (B’)

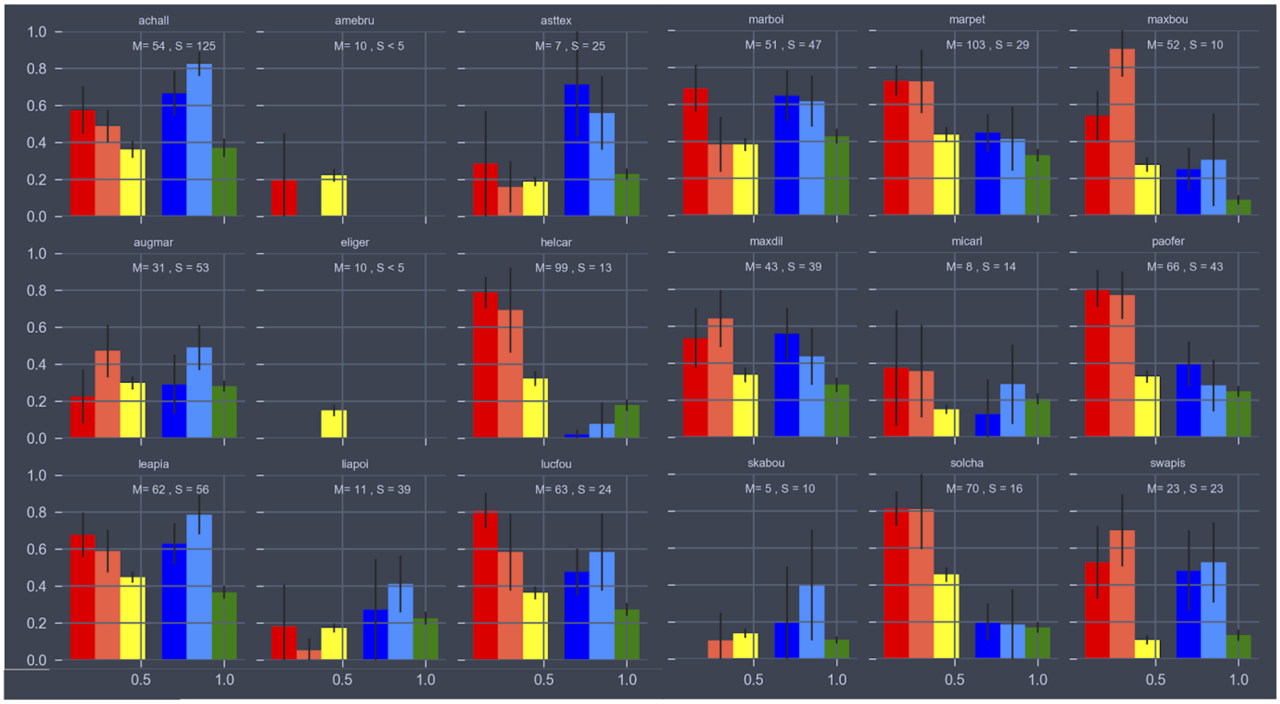

Figure 9 - P(B|B-St) (red), P(B|B’t) (tomato), P(B|random t) (yellow), P(B’|B-St) (dark blue), P(B’|B’t) (light blue) and P(B’|random t) (green) for each baby of the contingent group during the first 2 minutes with the highest number of occurrences of B’ within the period of test. The error bars (black) are the 95% confidence intervals for each value. M and S are the numbers of occurrences of B and B’ respectively from which P(B|B-St), P(B|B’t), P(B’|B-St) and P(B’|B’t) are computed.

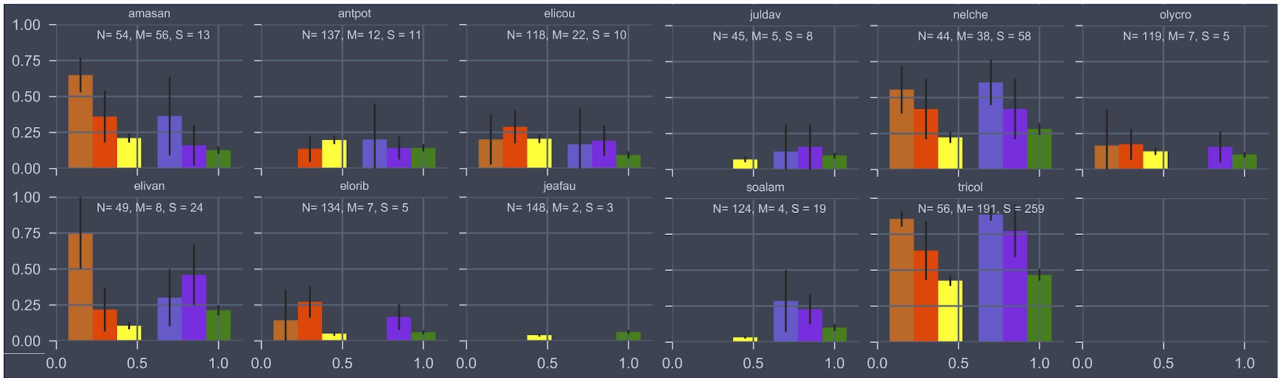

Figure 10 - P(B|Bt) (caramel), P(B|St) (blood orange), P(B|random t) (yellow), P(B’|B’t) (violet), P(B’|St) (flashy violet) and P(B’|random t) (green) for each baby in the control group during the first 2 minutes with the highest number of occurrences of B’ within the period of test. The error bars (black) are the 95% confidence intervals for each value. M and S are the numbers of occurrences of B and B’ respectively from which P(B|Bt), P(B|St), P(B’|B’t) and P(B’|St) are computed.

Appendix H. Second analysis in the 5-minute period of test

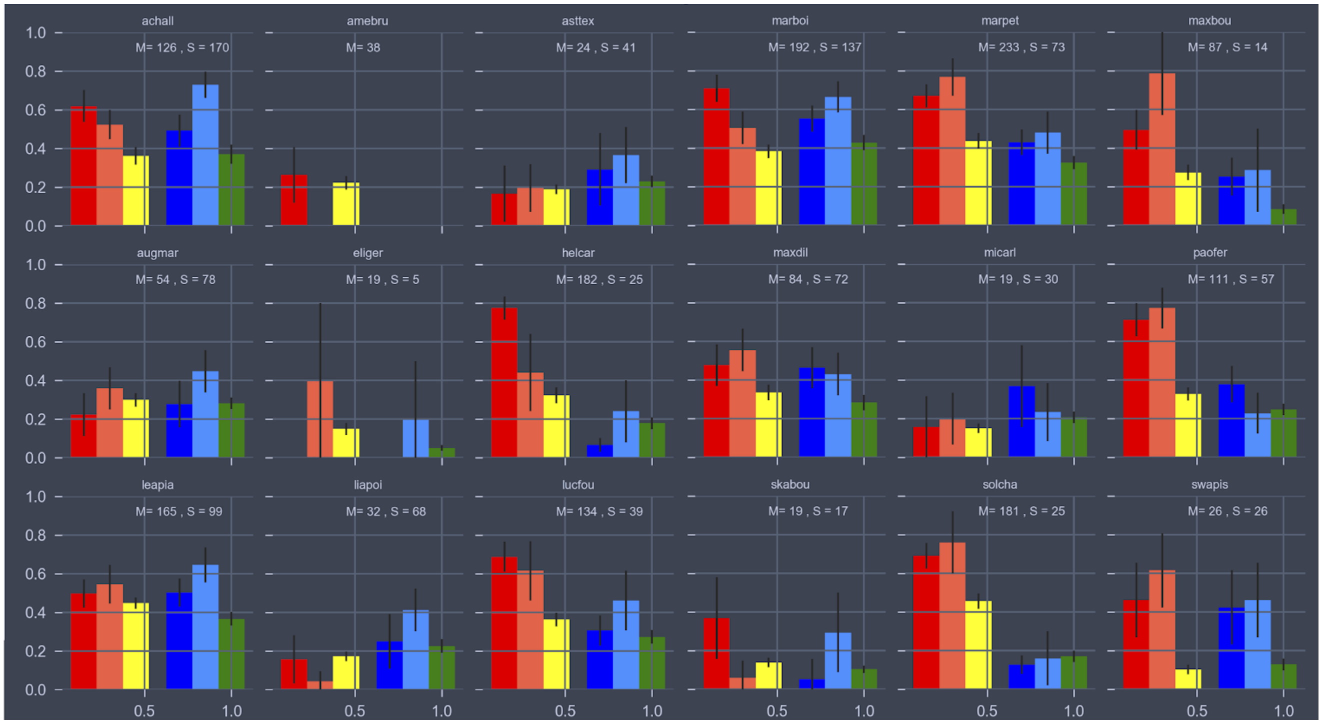

Figure 11 - P(B|B-St) (red), P(B|B't) (tomato), P(B|random t) (yellow), P(B'|B-St) (dark blue), P(B'|B't) (light blue) and P(B'|random t) (green) for each baby of the contingent group during the 5-minute period of test. The error bars (black) are the 95% confidence intervals for each value. M and S are the numbers of occurrences of B and B' respectively from which P(B|B-St), P(B|B't), P(B'|B-St) and P(B'|B't) are computed.

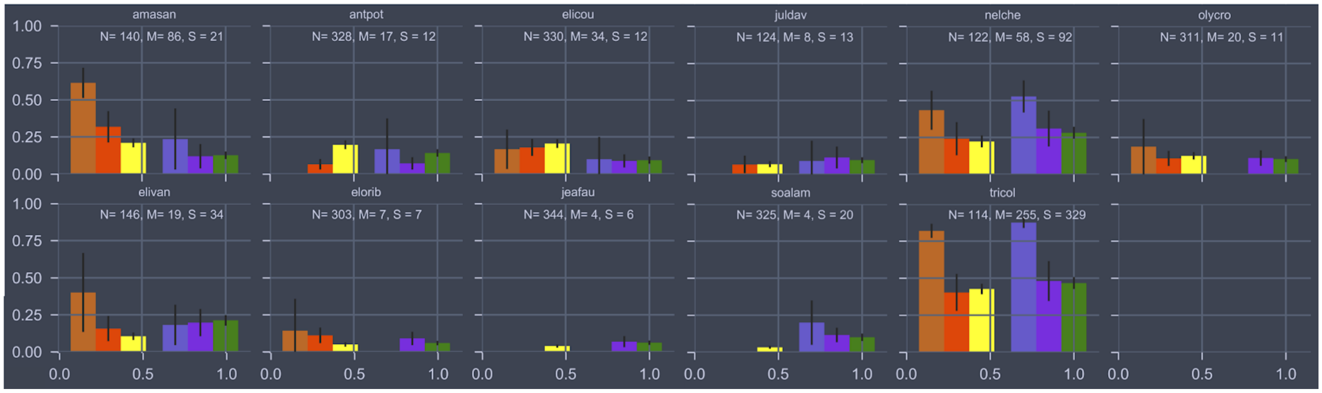

Figure 12 - P(B|Bt) (caramel), P(B|St) (blood orange), P(B|random t) (yellow), P(B'|B't) (violet), P(B'|St) (flashy violet) and P(B'|random t) (green) for each baby in the control group during the 5-minute period of test. The error bars (black) are the 95% confidence intervals for each value. M and S are the numbers of occurrences of B and B' respectively from which P(B|Bt), P(B|St), P(B'|B't) and P(B'|St) are computed.

Acknowledgements

Quatre mercis à toi Kevin : un pour avoir accepté de me rencontrer puis accepté de me revoir, un pour m'avoir présentée puis intégrée (et non imposée !) à ton équipe, un très gros pour les heures nombreuses que tu as passées à discuter, réfléchir, relire et corriger mon travail et surtout un grand merci pour les discussions scientifiques que tu nourris richement.

Merci à Sergiu pour ton humour fédérateur et l'aide précieuse que tu m'as apportée dans les analyses.

Merci Mark d'avoir réfléchi à nos problématiques et de nous avoir proposé des pistes.

Merci à Anastasia, Judith et Yasmine de m'avoir intégrée à leur bureau et un autre merci à Judith et Yasmine pour les samedis passés au labo dans la bonne humeur.

Merci à Lisa pour la littérature et les réponses à mes questions même pendant tes vacances.

Merci aux familles qui ont participé à notre étude.