CHAPTER 4

Over-estimation in numerosity production tasks: more than an attentional bias?[1]

Abstract

Over- and underestimation have been observed in numerosity estimation and approximate arithmetic tasks. Two different models have been proposed to account for these reverse patterns of performance. In estimation tasks, over- and underestimation have been interpreted as a consequence of a bi-directional mapping process between symbolic and non-symbolic representations of numbers (the symbolic to non-symbolic numerical mapping process leading to overestimation, and the non-symbolic to symbolic numerical mapping process leading to underestimation) (Castronovo & Seron, 2007a; Crollen, Castronovo & Seron, 2011). In approximate arithmetic tasks, over- and underestimation have been interpreted as reflecting an operational momentum, namely an attentional displacement along the mental number line (a rightward displacement leading to overestimation and a leftward displacement to underestimation) (McCrink et al., 2007). Our study was designed to examine whether the operational momentum could account for the overestimation found in numerosity production tasks. To this aim, 17 participants took part in a numerosity production task involving displacements along the mental number line. The results clearly demonstrated that overestimation was not merely due to an attentional bias, but rather relied on the cognitive operation of mapping two differently scaled numerical representations.

[1] This section is a modified version of an article with an identical title by Crollen and Seron, which has been submitted for publication to Acta Psychologica. The introduction section has been shortened in order to avoid redundancy with previous sections.

1. Introduction

Over- and underestimation errors were often observed in numerosity estimation and approximate arithmetic tasks, with the direction of the error depending on the numerosity estimation task used (i.e., production or perception) or the arithmetic operation (i.e., addition or subtraction) being carried out. Two different explanations, both using the metaphor of the mental number line as a theoretical background, have been proposed to account for these reverse patterns of performance: the bi-directional mapping hypothesis and the operational momentum hypothesis.

On the one hand and in line with the triple-code model, two recent studies (Castronovo & Seron, 2007a; Crollen et al., 2011) suggested that the over- and underestimation observed in numerosity estimation tasks were due to a bi-directional mapping/transcoding process between a non-symbolic numerical representation (i.e. the mental number line which is assumed to be logarithmically compressed and approximate) and a symbolic numerical representation which is assumed to be linear and precise.

On the other hand, over- and underestimation were also observed in approximate arithmetic tasks. More specifically, in a recent study, McCrink et al. (2007) showed adults videos of sets of objects being added or subtracted from one another and then required their participants to judge whether or not the final numerosity was the correct outcome of the operation. Overall, participants were able to confirm that the correct sum/difference of the two operands was indeed correct. More importantly, however, participants' responses showed a systematic bias (labeled the *operational momentum*): the outcomes of addition problems were systematically over-estimated, whereas the outcomes of subtraction problems were systematically under-estimated.

McCrink et al. (2007) postulated that these opposing patterns of performance were due to the numerical equivalent of a perceptual phenomenon called representational momentum. In a classic example of representational momentum, participants are shown a moving object that suddenly disappears. When they are asked to report the final position of this target object, they tend to displace the position in the direction of the implied motion: to the right with a rightward motion, and to the left with a leftward motion (Ashida, 2004; Freyd & Finke, 1984). According to McCrink et al. (2007), the parallels between representational and operational momentums are obvious: in the same way as moving objects present physical momentum, approximate arithmetic involves some displacements along a particular spatial frame, namely the left-to-right oriented mental number line. As small numbers are represented on the left of the number line and large numbers on the right, addition problems involve a rightward displacement on the mental number line, whereas subtraction problems yield a leftward displacement. So, when participants have to decide if a particular numerosity is or is not the correct outcome of an operation, their responses appear to be dependent on the arithmetic operation which was being carried out: the outcomes are overestimated for addition problems and underestimated for subtraction problems, indicating (analogously to the representational momentum) that participants are moving too far on their mental number line. This operational momentum has been replicated with children (McCrink & Wynn, 2009) and, to a lesser extent, with Arabic number symbols (Knops et al., 2009; Pinhas & Fischer, 2008). Moreover, a recent fMRI study (Knops et al., 2009) demonstrated that cortical maps for saccadic eye movement were activated during addition and subtraction, supporting the operational momentum hypothesis according to which an attentional displacement is evoked during approximate arithmetic.

Considering that participants realize their numerosity production judgments in an iterative way (e.g., production of successive key presses; Castronovo & Seron, 2007a; Cordes et al., 2001; Whalen et al., 1999), one may postulate that the numerosity production tasks actually involve a left-to-right displacement on the mental number line (the more the participants produce successive responses, the more they move towards the right of their mental number line). Unlike the discrete stimuli used by McCrink et al. (2007) in their approximate arithmetic tasks, the response mode required in the numerosity production tasks always exhibited continuous displacement. However, as representational momentum has been observed with discrete (Freyd & Finke, 1984, 1985) as well as with continuous stimuli (Hubbard & Bharucha, 1988), it is possible that participants overestimate numerosities in a numerosity production task because they move too far on their mental number line when producing their responses.

The goal of our study was therefore to examine whether the operational momentum hypothesis (McCrink et al., 2007) could be considered as an alternative interpretation to the overestimation error observed in numerosity production tasks. To this aim, 17 participants were required to produce the numerosity corresponding to a target Arabic numeral either by adding or by subtracting the number of dots presented in a starting display. If participants' performances are essentially driven by spatial displacements along the mental number line (the operational momentum hypothesis), then, participants should overestimate the numerosity when they are required to add dots to the starting display (make a rightward displacement on their mental number line) and, on the contrary, they should underestimate the numerosity when they are required to subtract dots from the starting display (make a leftward displacement on their mental number line). Alternatively, if participants' performances are driven by

the cognitive process of mapping the symbolic and non-symbolic representations of numbers (the bi-directional mapping hypothesis), then participants should overestimate the numerosity, irrespective of the direction of the attentional displacement to be done on the mental number line.

2. Method

2.1 Participants

The participants were 17 volunteers (11 females, 6 males) from the Catholic University of Louvain (Louvain-la-Neuve, Belgium) who were paid 10 euros each for their participation. They were aged between 18 and 21 (M = 19.35 ± 1.17), and were all right-handed.

2.2 Stimuli and procedure

The participants had to produce a dot pattern corresponding to the magnitude of a target Arabic numeral presented as input (30, 34, 36, 40, 44, 49, 52, 58, 64, 70, 78, 84, 91, 98).

As the use of dots implied the control of low-level continuous variables such as total occupied area, luminance, dot size and density (see Dehaene et al., 2005), two different sets of stimuli were created using Matlab: the Extensive set and the Intensive set. In the Extensive set, the sum of the area of all the dots (the luminance) and the total occupied area on the screen were kept constant across numerosities. As a consequence, the density (i.e., free area around each dot, comprising between 2 and 3 times the dot's radius) of the array increased, and the size of the dots decreased with increasing numerosity. In the Intensive set, the size of the dots and the density of the pattern were kept constant. As a consequence, the luminance and the total occupied

area increased with increasing numerosity. In both sets, six different images of each numerosity were created.

Each trial began with the presentation of a fixation cross for 500 ms, after which the target Arabic numeral appeared on the screen for 1000 ms. Then, a dot pattern was presented, indicating to the participants that they could start producing dots by using a potentiometer on a response box. The initial display contained either 7 (Rightward condition) or 254 (Leftward condition) dots, but no information was provided to participants about the numerosity of these displays (see Figure 4-1).

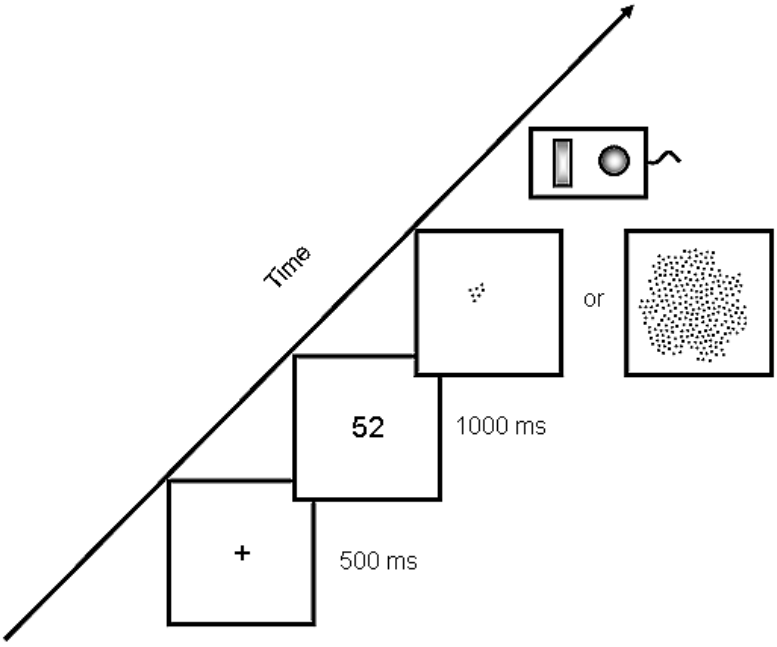


Fig. 4-1. Illustration of the numerosity production task's method

Previous studies have found that adults are able to discriminate between arrays that differ by a ratio of .87 − .90 (Halberda & Feigenson, 2008; Pica et al., 2004; Temple & Posner, 1998; van Oeffelen & Vos, 1982). Here, the starting array with 254 dots differed from the largest target numerosity (98) by a ratio of .38, and the

starting array with 7 dots differed from the smallest target numerosity (30) by a ratio of .23. So participants should certainly have realized that, for each trial in the Rightward condition they had to increase the number of dots on the computer screen (i.e. make a rightward displacement on their mental number line) by turning the potentiometer to the right, and vice versa in the Leftward condition. Moreover, the experimenter emphasized that the biggest display (254) would always contain too many dots, and the smallest (7) would always contain too few dots.

Due to technical constrains, participants were not able to produce more than 254 dots. In order to avoid the use of an overt or covert counting strategy, the dots appeared quickly on the screen (creation of one dot every 1.4° of angular rotation) and participants were explicitly instructed not to count the number of dots they produced. They were also required to perform the task as fast as possible and had to report their completion by pressing a response button on the response box.

Each target number was presented 24 times into 12 experimental blocks of 28 trials. The Arabic numerals in each block were randomly presented (2 presentations of the same number per block – one inducing the production of an Extensive stimulus and one inducing the production of an Intensive stimulus). The whole session lasted approximately an hour. In half of the blocks, the Rightward condition was associated with the Extensive set and the Leftward condition with the Intensive set, and in the other half, the reverse was true. The order of the blocks was counterbalanced across participants, so that half of them worked from Block 1 to Block 12, and the other half in the reverse direction. The blocks were separated by pauses when required by the participants. Prior to the experiment, the participants were given 8 training trials. Moreover, participants received no information about the range of numerosities to be presented and no feedback regarding their performance.

3. Results

Data from trials on which a participant's response was more than three standard deviations greater than the participant's overall mean were excluded from the analyses (0.9% of the results). Moreover, as preliminary analyses demonstrated that the mean estimates and their standard deviations were not normally distributed, the following analyses were conducted on their logarithmic transformations.

LMM with numerosity as a fixed effect and subject as a random effect were used to examine whether participants' patterns of performance obeyed Weber's law by showing scalar variability (i.e., constant coefficients of variation). In both conditions, the log(mean number of dots produced) increased with the target magnitude, $F(13, 224) = 34.16$, $p < .001$ for the Rightward condition; $F(13, 224) = 45.03$, $p < .001$ for the Leftward condition, as did the log(standard deviation), $F(13, 224) = 11.28$, $p < .001$ for the Rightward condition; $F(13, 224) = 5.97$, $p < .001$ for the Leftward condition.

The coefficients of variation, CV_s, were then computed and analyzed using a LMM. In both conditions, CV_s were constant, $F(13, 224) = .72$, $p > .7$ in the Rightward Condition; $F(13, 224) = .91$, $p > .5$ in the Leftward Condition, indicating that the scalar variability principle provided an adequate description of the data.

Finally and more importantly, t-tests with a test value of zero were conducted on the response bias, RB ((participant's response − target magnitude)/target magnitude) to see whether the participants' responses were accurate, underestimated or overestimated. As a preliminary paired-sample t-test showed that the RB_s were similar in the Extensive ($M = .72 \pm .45$) and Intensive ($M = .77 \pm .34$) sets, $p > .5$, data from the two stimuli sets were merged. The RB would be zero if the estimates were accurate. A negative RB indicates that magnitudes were underestimated,

whereas a positive RB indicates that magnitudes were overestimated. Participants overestimated target magnitude, in the Rightward Condition (M = .58 ± .38), $t(16)$ = 6.21, p < .001, as well as in the Leftward condition (M = .91 ± .34), $t(16)$ = 11.06, p < .001. This response bias was observed on every target magnitude (see Figure 4-2), all p_s < .001. However, the responses were more overestimated in the Leftward condition than in the Rightward condition, $t(16)$ = −5.94, p < .001.

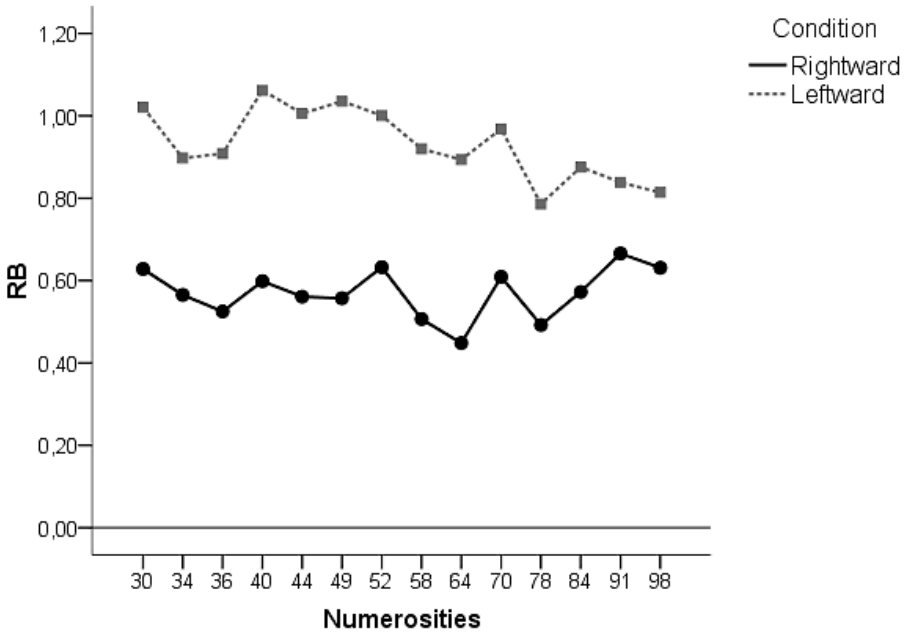


Fig. 4-2. Response bias (RB) in the production task: participants overestimated every target numerosity, in the Rightward as well as in the Leftward condition.

4. General discussion

The present results confirmed that the participants' behavior followed the well-established psychophysical Weber's law. Indeed, the means and the standard deviations of the numerical judgments increased with the target magnitudes and in direct proportion. This resulted in constant coefficients of variation across target size

(Castronovo & Seron, 2007a; Cordes et al., 2001; Whalen et al., 1999), and indicated that the scalar variability principle well described our data.

Moreover and more importantly for our purposes, our data also showed that the general tendency to overestimate numerosities in a numerosity production task may not be merely due to attentional processes that mimic displacements along the mental number line (i.e., operational momentum). Indeed, although the production task used involved a leftward as well as a rightward displacement on the mental number line, overestimation occurred in both conditions (not overestimation in the rightward condition and underestimation in the leftward condition as predicted by the operational momentum hypothesis). Therefore, our study gave further support to the idea that overestimation originates from the numerical mapping process that takes place between the symbolic and non-symbolic representations of numbers.

Participants' estimates were nevertheless influenced by the numerosity supplied as a starting point for the estimation process (7 or 254). However, while the operational momentum hypothesis would have predicted larger responses in the Rightward condition than in the Leftward condition, our data showed exactly the opposite: target magnitudes were more overestimated in the Leftward than in the Rightward condition. This difference between the two conditions might be explained by an anchoring effect, which is the tendency to rely on specific information or specific value (i.e., the anchor) during normal decision making (Tversky & Kahneman, 1974). This cognitive bias has already been shown to influence the estimation of weights (DiLollo, 1964), emotional expressions (Manis & Armstrong, 1971), percentages (Tversky & Kahneman, 1974) and historical dates (Strack & Mussweiler, 1997). In all cases, the influence is towards the anchor. In the present study, the participants had to produce dot patterns by starting from an initial display

which contained either 7 or 254 dots. It is possible that participants took this starting display as an anchor, and then adjusted their production judgements upward or downward from this anchor until they reached a "plausible" answer (see, for example, Tversky & Kahneman, 1974). Numerosities which were further away from the anchor might also be plausible, but participants stopped estimating at the first answer they felt was satisfactory, so inducing a larger overestimation error in the Leftward than in the Rightward condition.

One might wonder whether the response mode used in our task is sufficient to induce attentional displacements on an internal number line. Indeed, the fact that our task involves a displacement in the way participants have to produce their "external" response does not necessarily mean that our tasks involve a displacement on an internal representation. However, it is worth nothing that all the primary studies examining representational momentum in the visual field actually focused on the parallels between the external implied motion of the target and the internal representation of this target. Finke & Freyd (1985; see also Finke, Freyd, & Shyi, 1986), for example, suggested that shifts in visual memory occur because there is a natural tendency to mentally extrapolate the motion of a target stimulus. Just like a moving physical object, this extrapolation process presents a kind of inertia and, as a consequence, the mental extrapolation cannot be stopped immediately but will continue to coast some distance after a force has been applied to stop it. Similarly, in our production task, it is possible that participants do not immediately stop the numerosity estimation process because the implied dynamics of the required response mode possess its own momentum and therefore lead participants to coarse some more distance on their internal representation of numbers.

To sum up, our study clearly demonstrated that the overestimation found in numerosity production tasks cannot be attributed solely to spatial displacements on the non-symbolic representation of numbers. Instead, our findings strengthen the bi-directional mapping hypothesis's postulate according to which overestimation depends on the mapping process that takes place between two numerical representations which are scaled in different ways. By pre-exposing participants to two different stimuli (7 and 254), our experimental design nevertheless influenced participants' patterns of performance. Importantly, however, this observation can not be interpreted as reflecting any kind of operational momentum.

CHAPTER 5

Calibrating the numerosity estimation process reflects the differential coding of symbolic and non-symbolic representations of numbers[2].

Abstract

A recent computational model suggested that small symbolic and non-symbolic quantities were processed by distinct representational systems before activating a shared numerical representation, the symbolic numbers undergoing a place-coding process and the non-symbolic quantities undergoing a summation-coding step (Roggeman et al., 2007; Verguts & Fias, 2004). Using numerosity perception and production tasks, the present study aimed to test whether the calibration of the estimation process could reflect these representational differences in the range of large numerical values. Feedback trial and participants' response format were manipulated (Arabic digits in the perception task and collection of dots in the production task) as well as the distance between the feedback trial and the target magnitude to estimate. The results of our experiments revealed that the shape of the calibration pattern was dependent on the format used to present the feedback trial and the participants' responses. When both were presented in the symbolic code (i.e., Arabic digits), the calibration effect reflected the characteristics of the place-coded representation. Conversely, when both were presented in non-symbolic notation (i.e., collection of dots), the calibration effect reflected the characteristics of the summation-coded representation. These results therefore demonstrated, at a

[2] This section is a modified version of an article with an identical title by Crollen and Seron, which has been submitted for publication to Acta Psychologica. The introduction was shortened in order to avoid redundancy with the previous sections.

behavioural level, that the processing of large numerical values was supported by different underlying representations.

1. Introduction

Recent studies supported the idea that a shared numerical representation is reached through different pathways, depending on the stimlus format: whereas non-symbolic quantities reach the representation of numbers through a summation coding-step, the processing of symbolic quantities activated a pathway that does not rely on this intermediate stage (Roggeman et al., 2007; Santens et al., 2010; Verguts & Fias, 2004).

Besides their differential coding, symbolic and non-symbolic representations of number also present hemispheric asymmetry (e.g., Ansari et al., 2005; Holloway et al., 2010; Piazza et al., 2007) and a different kind of scaling (Dehaene, 2009), the symbolic representation being linearly scaled (Verguts et al., 2005) and the non-symbolic representation logarithmically compressed (Dehaene, 1992; Verguts & Fias, 2004). A linear scaling means that the distance between numbers on the number line is equal for all the numbers with a fixed distance (Verguts et al., 2005; Booth & Siegler, 2006). A logarithmic scaling, on the other hand, means that the distance between numbers on the number line is smaller for larger numbers (Brysbaert, 1995; Dehaene, 1992; Verguts & Fias, 2004). Therefore, when these two differently scaled representations are mapped to each other, the magnitude reached on the symbolic representation of numbers is always smaller than the magnitude reached on the non-symbolic representation. Crollen and colleagues (2011) recently demonstrated how this mapping mechanism could be the origin of the under- and overestimation error observed in numerosity estimation tasks (see also Castronovo & Seron, 2007a).

It is important to note that the general tendency to underestimate numerosity in a numerosity perception task can be calibrated if symbolic feedback is provided to participants (Izard & Dehaene, 2008; Minturn & Reese, 1951).

To date, however, we do not know whether it is possible to calibrate the numerosity production process by giving a non-symbolic feedback to participants. The first aim of the present study is therefore to examine this issue. By doing this, we will also be able to determine whether the feedback and response format could have an impact on the the calibration process. It is indeed possible that calibration occurs at every numerosity but with a different strength depending on the feedback and response format and on the distance separating the feedback and the target magnitudes. If both the feedback trial and the participants' response are presented as Arabic digits, as in the two previous studies (Izard & Dehaene, 2008; Minturn & Reese, 1951), the calibration effect could reflect the place-coded feature of the symbolic representation of numbers. Indeed, by activating a delineated segment of the symbolic representation of numbers around the feedback value, symbolic feedback trial could improve the non-symbolic to symbolic numerical mapping by more when the target quantity falls within this segment than when it does not. Therefore, in the same way as a symbolic prime facilitated the processing of symbolic targets smaller and larger than but close to the prime value (Roggeman et al., 2007), a symbolic feedback could induce a distance-related calibration effect on the symbolic responses smaller and larger than the feedback value. Conversely, if both the feedback and the response are presented as collections of dots, the calibration effect could reflect the summation coded characteristic of the non-symbolic numerical representation. Indeed, as the non-symbolic feedback will activate not only the feedback value but

also all smaller numerosities, it is possible that the calibration effect will be larger for this numerical range than for the numerosities larger than the feedback value.

To our knowledge, however, two non-symbolic priming studies did not reveal priming at the summation-coding stage (Di Luca et al., 2010; Herrera, & Macizo, 2008). Herrera and Macizo (2008), on the one hand, examined the semantic priming between sets of dots (random or canonical sets) and Arabic digits in a numerical comparison task and found cross-notational semantic priming: when dots were presented in canonical patterns, semantic priming was found in both cross-notational conditions, altough it was weaker for digit-dice combinations. When dots were presented in random configuations, semantic priming only occurred from dots to digits. Di luca et al. (2010), on the other hand, revealed that non-symbolic canonical finger configurations could induce priming effect of the place-coding type. Contrary to what was observed by Roggeman et al. (2007), these studies did not find specific curves related to prime notation; and therefore suggest that a summation coding is not all that evident for non-symbolic quantities. According to us, these apparent diverging results could nevertheless be explained by two different factors. The first factor is related to the ease of relating the non symbolic numerosity to an exact cardinal meaning. Because canonical configurations of dots (Herrera & Macizo, 2008) and canonical configurations of fingers (Di Luca et al., 2010) are familiar and conventional signs, this kind of non-symbolic stimuli should be directly associated with an exact value and hence entail a magnitude representation similar to that produced by symbolic notation (i.e., V-shaped). The second factor is related to the mapping process that is required to attach a symbolic label to a particular quantity. While this process might be absent in comparison task (Herrera & Macizo, 2008), it is essential in naming task (Roggeman et al., 2007).

As the perception and production tasks usually involve random sets of dots and a numerical mapping process, the study of the calibration effect could be a good way to examine wether the place-coding and summation-coding steps characterize the processing of large numerical values, as well as the processing of small numerical values (Roggeman et al., 2007).

In this paper, two groups of 18 adults took therefore part in a numerosity perception task or in a numerosity production task. The numerosity perception task required the estimation of the number of dots presented within a display. In the first half of the experiment, participants had to give their responses without having any information about the numerical range tested. In the second half, participants were explicitly told that a given pattern contained 60 dots and were then required to take this feedback trial into account to produce their subsequent responses. If the calibration effect reflects the place-coding principle, then the performance improvement should peak at the feedback value but should also be observed at smaller and larger neighboring values, with gradually decreasing strength. Conversely, if the calibration effect is not related to the particular coding of the symbolic numerical representation, then the performance improvement should be the same, irrespective of the distance separating the feedback value and the target magnitude.

In the numerosity production task, participants were required to produce a dot pattern corresponding to a target Arabic numeral. Spontaneous responses were recorded in the first half of the experiment. In the second half, responses were recorded after a pattern of 60 dots were explicitly shown to the participants. If the calibration effect reflects the summation-coding principle, then the feedback trial should improve the estimation of the numerosities smaller or equal to the feedback

value by more than the estimation of the numerosities larger than the feedback value. Conversely, if the calibration effect does not reflect the summation-coding feature of the non-symbolic numerical representation, then the calibration measure should be the same, irrespective of the value of the target magnitude.

2. Experiment 1: numerosity perception task

2.1 Participants

The participants were 18 volunteers (16 females, 2 males) from the "Université Catholique de Louvain" (Louvain-la-Neuve, Belgium) who received course credit for their participation. The average age was 20.8 (range: 19−30 years; *SD* = 2.96). Two participants were left-handed.

2.2 Stimuli and procedure

In this task, fifteen numerosities ranging from 25 to 95 were used (i.e., 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95) to create dot patterns. The numerical range started at 25, as previous studies showed that numerical estimates are usually accurate for numerosities smaller than 20 (Indow & Ida, 1977; Krueger, 1982, 1984). Participants had to estimate the numerosity of the display, by using Arabic numerals. To prevent participants from basing their responses on low-level continuous variables such as total occupied area, luminance, dot size and density (see Dehaene et al., 2005), two different sets of stimuli were created by using Matlab. In the first set (i.e., the Extensive Set), the sum of the area of all the dots (the luminance) and the total occupied area on the screen were kept constant across all numerosities. Thus, in this set, the density of the array increased with increasing numerosity and the size of the dots decreased with increasing numerosity. In the second set (i.e., the Intensive Set),

in contrast, the size of the dots as well as the density of the pattern were kept constant. As a consequence, the luminance and the total occupied area increased with increasing numerosities. In both sets, six different images of each numerosity were created. The stimuli were black on a white background.

Each trial began with the presentation of a fixation cross (500 ms) after which a dot pattern was flashed on the screen for a duration of 150 ms. Participants typed their responses in Arabic format on the computer number pad, and then pressed the space bar to validate their answer. The experiment consisted of two sessions, each containing 6 blocks of 30 trials. In the first six blocks (i.e., the pre-feedback session), participants estimated the numerosity of the dot patterns without having any indication of the numerical range tested. In the last six blocks (i.e., the post-feedback session), participants received feedback on their performance: for the first item of each block, the expected Arabic number was presented to participants following the production of their responses. The feedback trial followed either the presentation of an Extensive dot pattern (three blocks) or the presentation of an Intensive stimulus (three blocks). The feedback value was, in all cases, fixed at 60 because this value is in the middle of the numerical range tested. Therefore, this allowed us to calculate the calibration effect similarly for the numerosities smaller and larger than the feedback value.

Each target numerosity was presented 24 times (2 presentations of the same numerosity in each block – one from the Extensive Set and one from the Intensive Set), in a random order. The order of the blocks was counterbalanced across participants. Half the participants undertook Block 1 to Block 12, and the other half Block 12 to Block 1. The blocks were separated by pauses when required by the

participants. Prior to the experimental trials, participants were given six training trials.

2.3 Results

Data for the trials in which a participant's response was three standard deviations greater than the participant's overall mean response were excluded from the analysis. Moreover, as the mean estimates and their standard deviations were not normally distributed, the following analyses were conducted on their logarithmic transformation.

LMM with numerosity as a fixed effect and subjects as a random effect were conducted to determine whether participants' patterns of performance obeyed Weber's law by showing scalar variability (i.e., constant CVs). In both sessions, the log of the mean estimates increased linearly with the target magnitude, $F(14, 255) = 23.07$, $p < .001$ in the pre-feedback session; $F(14, 255) = 26.03$, $p < .001$ in the post-feedback session, as did the log of the standard deviation, $F(14, 255) = 6.57$, $p < .001$ in the pre-feedback session; $F(14, 255) = 3.43$, $p < .001$ in the post-feedback session. The coefficients of variation were then computed (CV = standard deviation/mean) and analyzed using LMM. The results were not statistically significant suggesting that CV_s did not vary with numerosity and were constant across target size, $F(14, 255) = 1.03$, $p > .4$ in the pre-feedback session; $F(14,255) = 0.61$, $p > .8$ in the post-feedback session.

To see whether the participants' responses were accurate, underestimated or overestimated, t-tests with a test value of 0 were conducted on the error rate (i.e., (participant's response – target magnitude)/target magnitude) of each session. The error rate (ER) should be zero if estimates are accurate. A negative ER would indicate

that magnitudes were underestimated, whereas a positive ER would indicate that magnitudes were overestimated. The results indicated that participants' responses were underestimated in the pre-feedback session, $M = -.42 \pm .18$, $t(17) = -10.20$, $p < .001$, as well as in the post-feedback session, $M = -.16 \pm .19$, $t(17) = -3.60$, $p < .01$ (see Fig. 5-1).

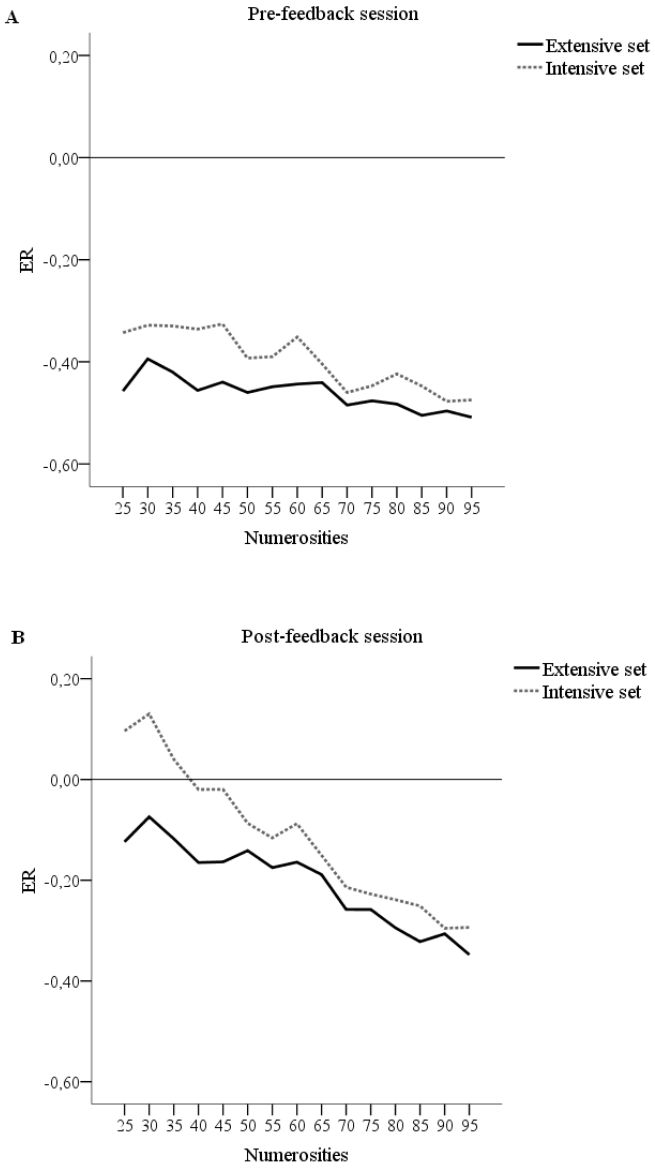


Fig. 5-1. ER for each stimulus set of the numerosity perception task in **(A)** the pre-feedback session and **(B)** the post-feedback session.

A Session (Pre-Feedback, Post-Feedback) x Stimuli Sets (Extensive, Intensive) repeated measures ANOVA was then performed on the ER in order to examine the effects of these two factors on participants' performance. This analysis showed: **(1)** a main effect of the Session, $F(1, 17) = 41.10$, $p < .001$, reflecting that ER was more important in the pre-feedback session ($M = -.42 \pm .18$) than in the post-feedback session ($M = -.16 \pm .19$) (see Fig. 5-1); **(2)** a main effect of the Stimuli Sets, $F(1, 17) = 35.24$, $p < .001$, showing that responses were, overall, more underestimated in the Extensive set ($M = -.33 \pm .29$) than in the Intensive set ($M = -.25 \pm .29$). The Session x Stimuli Sets interaction was also significant, $F(1, 17) = 5.14$, $p < .05$, indicating that the difference between the two stimuli sets was more pronounced in the post-feedback session ($M = -.09 \pm .07$) than in the pre-feedback session ($M = -.06 \pm .05$) (see Fig 5-1).

Finally, and most importantly for our purposes, a calibration measure (i.e., $|\text{ER pre-feedback}| - |\text{ER post-feedback}|$) was calculated in order to examine the effect of the feedback trial. In order to examine the shape of the calibration curve, we fitted regression equations with two predictors that coded for a step-function and an inverted V-function, respectively (see Roggeman et al., 2007). The step-function predictor had a coefficient of −1 if target value was smaller or equal to feedback value and a coefficient of +1 if target value was larger than feedback value. A negative regression coefficient for the step function predictor indicates that the shape of the calibration curve can be described by a step-function in which target values smaller or equal to the feedback value lead to larger calibration effect. In contrast, a negative coefficient for the inverted V-function predictor indicates that the shape of the calibration curve can be described by an inverted V-function in which smaller $|\text{target value} - \text{feedback value}|$ distances lead to larger calibration effect. The

regression was run for each participant separately (Lorch & Myers, 1990) and the coefficients then obtained were submitted to a one-sample t-test with a test value of 0: the inverted V-function predictor was different from 0 ($M = -.34$), $t(17) = -3.88$, $p = .001$, while the step-function predictor was not ($M = -.03$), $t(17) = -.22$, $p > .8$ (see Fig. 5-2). These results therefore indicated that the inverted V-function was the best predictor for the calibration effect of the perception task.

As a mere dichotomic categorization between targets smaller and larger than primes values may be insufficient to correctly account for the data (see Di Luca et al., 2010), the calibration measure was submitted to a LMM with participants as a random factor, the valence of the regression distance (negative for the target numerosities smaller than the feedback value and positive for the target numerosities larger than the feedback value) as a fixed factor and the absolute distance between the feedback and the target numerosity (5, 10, 15, 20, 25, 30, 35) as a covariate. The calibration measure should be zero if the feedback has no effect. A negative calibration measure would indicate that the feedback impaired participants' performances, whereas a positive calibration measure would indicate that the feedback improved participants' performances. The analysis demonstrated that the calibration effect was shaped as an inverted V. Indeed, the distance effect was significant, $F(6, 238) = 2.49$, $p < .05$, whereas the valence and the interaction between the distance and the valence were not, $F(1, 238) = 1.69$, $p > .1$ for the valence factor; $F(6, 238) = 0.42$, $p > .8$ for the distance x valence interaction. Therefore, and as shown in Figure 5-3, the symbolic feedback provided in the perception task appeared to improve participants' performance (i.e., positive calibration measure). This improvement was inversely proportional to the distance separating the feedback and the target numerosity (i.e., distance effect: the more

distant from the feedback was the target numerosity, the less important was the calibration) and occurred with numerosities smaller and larger than the feedback value (i.e., no valence effect).

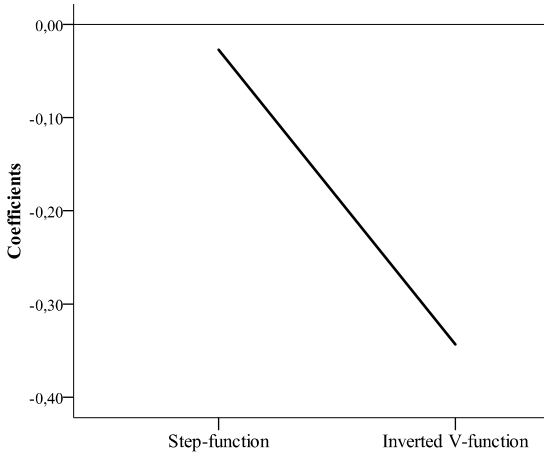


Fig. 5-2. Regression coefficients for the predictors describing step-like and inverted V-shape calibration functions of the perception task.

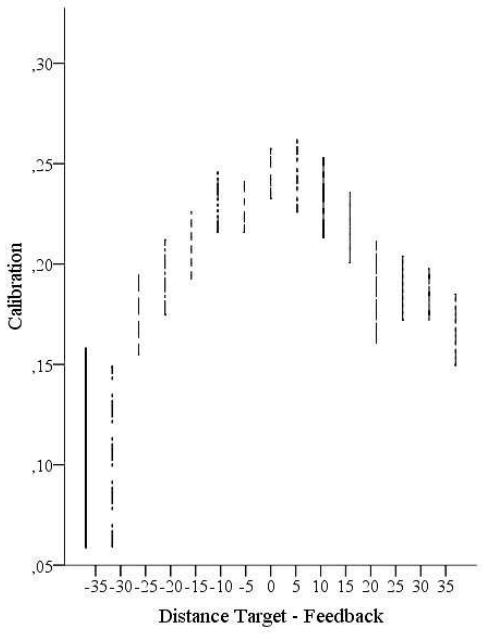


Fig. 5-3. Inverted V-shaped calibration effect observed in the perception task (with a symbolic feedback trial and symbolic responses). Error bars were computed using Cousineau's method (2005). Cousineau's method allows the data of a repeated measures design to be represented without masking information such as the heterogeneity of variances across conditions. In this method, each participant's score is adjusted so that its relative position is no longer present.

2.4 Conclusions

Altogether, these first results replicated patterns of performance previously found in studies on numerical estimation: **(1)** both the means and the standard deviations of the numerical judgments increased with the target magnitudes and in direct proportion, resulting in constant coefficients of variation across target size (Castronovo & Seron, 2007a; Cordes et al., 2001; Whalen et al., 1999), and indicating that the scalar variability principle well described our data; **(2)** participants underestimated magnitudes in a numerosity perception task (e.g., Izard & Dehaene, 2008); **(3)** providing participants with some information about the numerosities tested improves the accuracy of the responses (Izard & Dehaene, 2008; Minturn & Reese, 1951); **(4)** estimates are modulated by non-numerical parameters. The finding that the Intensive dot arrays appeared to be more numerous than the Extensive dot arrays is consistent with Ginsburg's observation (1978, 1980) that arrays are judged to contain more dots when the dots are regularly spaced.

More importantly for our purposes, this first experiment also showed that the feedback effect was in line with what would be expected from a place-coded representation of numbers: the feedback trial improved numerosity perception judgments. This improvement peaked at the feedback value but also extended to neighbouring numerosities (smaller and larger than the feedback value), with a decreasing strength.

In order to examine whether non-symbolic feedback could improve numerosity production judgments in a way reflecting the summation-coding feature of the non-symbolic numerical representation, a second experiment was run in which participants were required to produce a dot pattern corresponding to a target Arabic

numeral. The production judgments were evaluated before and after the presentation of a non-symbolic feedback trial.

3. Experiment 2: numerosity production task

3.1 Participants

The participants were 18 volunteers (14 females, 4 males) from the "Université Catholique de Louvain" (Louvain-la-Neuve, Belgium) who received course credit for their participation. The average age was 21.16 (range: 19−28 years; *SD* = 2.68). One participant was left-handed.

3.2 Stimuli and procedure

In the numerosity production task, the numerical range (i.e., 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95) and materials (i.e., Extensive and Intensive patterns of dots) used were the same as in the numerosity perception task of Experiment 1.

Each trial began with the presentation of a fixation cross (1000 ms) after which the target Arabic numeral appeared on the screen for a duration of 1000 ms. After showing the target number, a single dot was presented, indicating to the participants that they could start production of the dots by turning a potentiometer on a response box (to the right to increase the number of dots appearing on the screen and to the left to decrease the number of dots). The response box allowed participants to produce all the numerosities ranging from 1 to 254. However, participants were encouraged to use as few back and forth movements as possible to get to their final response. Half of the pattern of dots produced by participants was initiated from the Extensive set and the other half from the Intensive set. In order to avoid the use of an overt or covert

counting strategy, the dots appeared quickly on the screen (creation of one dot for every 1.4° of angular rotation) and participants were explicitly instructed not to count the number of dots they produced. They were also required to perform the task as fast as possible and had to report their completion by pressing a response button on the response box.

The experiment consisted of two sessions, each containing 6 blocks of 30 trials. In the first six blocks (i.e., the pre-feedback session), participants produced the dot patterns without having any information about the accuracy of their responses. In the last six blocks (i.e., the post-feedback session), participants received feedback on their performance: for the first item of each block, the expected dot pattern was presented to participants following the production of their responses. In this experiment, the feedback trial consisted in the presentation of a collection of 60 dots and followed either the production of an Extensive dot pattern (three blocks) or the production of an Intensive stimulus (three blocks).

The presentation and number of practice and experimental trials were identical to those in Experiment 1.

3.3 Results

Data for the trials in which a participant's response was three standard deviations greater than the participant's overall mean response were excluded from the analysis. Moreover, as the mean estimates and their standard deviations were not normally distributed, the following analyses were conducted on their logarithmic transformation.

Linear mixed models (LMM) with numerosity as a fixed effect and participants as a random effect were then performed to determine whether participants' patterns of

performance obeyed Weber's law by showing scalar variability (i.e., constant CVs). In this task, the log of the mean estimates increased linearly with target magnitude, $F(14, 255) = 30.35$, $p < .001$ in the pre-feedback session; $F(14, 255) = 125.38$, $p < .001$ in the post-feedback session, as did the log of the standard deviation, $F(14, 255) = 14.76$, $p < .001$ in the pre-feedback session; $F(14, 255) = 31.75$, $p < .001$ in the post-feedback session. The CV_s were then computed and analyzed using LMM. The results were not significant, suggesting that CV_s did not vary with numerosity neither in the pre-feedback session, $F(14, 255) = 1.54$, $p = .1$, nor in the post-feedback session, $F(14, 255) = 0.71$, $p > .7$.

To see whether the participants' responses were accurate, underestimated or overestimated, t-tests with a test value of 0 were conducted on the ER of both sessions. As predicted, participants' responses were overestimated in the pre-feedback session, $M = .79 \pm .55$, $t(17) = 6.06$, $p < .001$, as well as in the post-feedback session, $M = .31 \pm .20$, $t(17) = 6.67$, $p < .001$ (see Fig. 5-4).

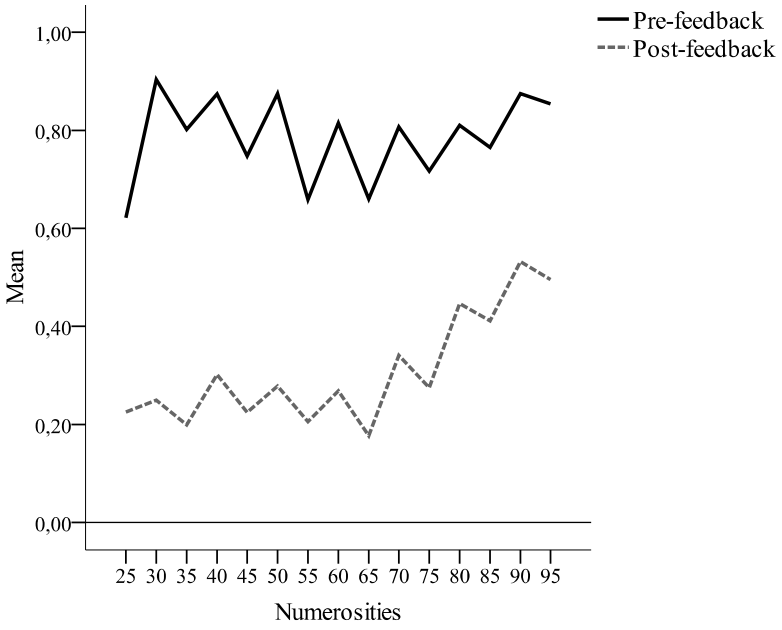


Fig. 5-4. ER of the numerosity production task in the pre- and post-feedback session.

Subsequently, a Session (Pre-feedback vs. Post-feedback) x Stimuli Sets (Extensive vs. Intensive) repeated measures ANOVA was run on the ER. This analysis showed a significant main effect of the Session factor, $F(1, 17) = 16.92$, $p < .01$, revealing that the ER was larger in the pre-feedback session ($M = .79 \pm .55$) than in the post feedback session ($M = .31 \pm .20$) (see Fig. 5-4). The stimuli sets effect, in contrast, was not significant, $F(1, 17) = 0.59$, $p > .4$, nor was the Stimuli Sets x Session interaction, $F(1, 17) = 0.25$, $p > .6$.

More importantly for our purposes, regression equations were calculated on the calibration measure fore each participant separately. As in the previous task, these regressions were fitted with two predictors (one coding the step-function and one coding the inverted V-function) and then submitted to a one-sample t-test with a test value of 0. In this production task, the step-function predictor was different from 0 ($M = -.30$), $t(17) = -3.32$, $p = .01$, while the inverted V-function predictor was not ($M = -.09$), $t(17) = -.22$, $p > .2$ (see Fig. 5-5). These results therefore indicated that the step-function was the best predictor for the calibration effect of the production task.

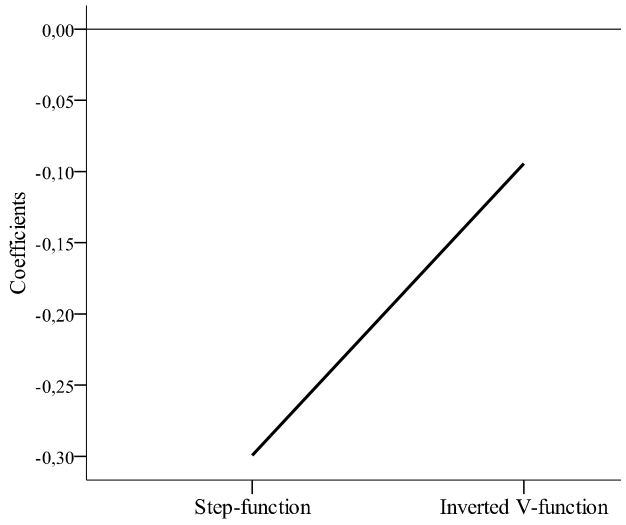


Fig. 5-5. Regression coefficients for the predictors describing step-like and inverted V-shape calibration functions of the production task

A LMM (with participants as a random factor, absolute distance as a covariate and valence as a fixed factor) was run on the calibration measure in order to further examine whether the calibration effect reflects the summation-coding feature of the non-symbolic numerical representation. As numerosities 55 and 65 were not easily distinguishable from the feedback value (because they differed by a ratio of 0.9), the analysis was run without the distance of 5. The results demonstrated that the valence effect was significant, $F(1, 204) = 4.77$, $p < .05$, whereas the distance effect and the distance x valence interaction were not, $F(5, 204) = 0.34$, $p > .8$ for the distance effect; $F(5, 204) = 0.37$, $p > .8$ for the distance x valence interaction. The main effect of valence indicated that the improvement in performance was more pronounced with numerosities smaller than the feedback value ($M = .53 \pm .52$) than with numerosities larger than the feedback value ($M = .40 \pm .47$). The calibration effect thus appeared to be shaped like a step-like function (see Figure 5-6).

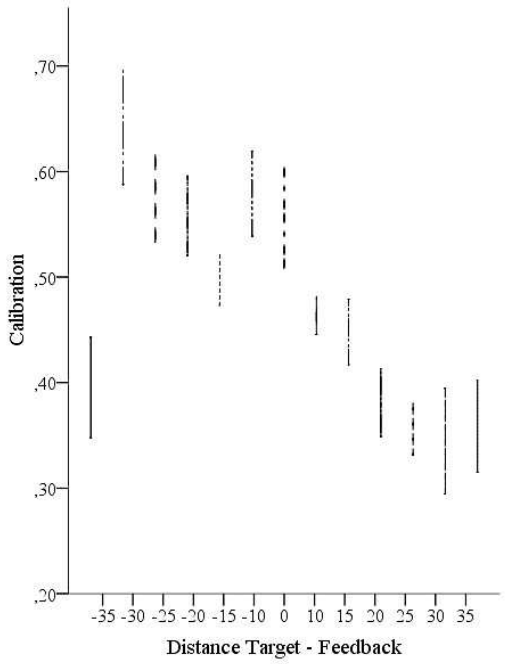


Fig. 5-6. Step-like calibration effect observed in the production task (with a non-symbolic feedback trial and non-symbolic responses). Error bars were computed using Cousineau's method (2005).

Moreover, it is interesting to note that the non-symbolic feedback of the numerosity production task leads to a greater calibration effect (M = .48 ± .49) than the symbolic feedback provided in the numerosity perception task (M = .23 ± .12), $t(34) = -2.15$, $p < .05$. This probably results from the fact that participants' absolute error rate was originally larger in the production task (M = .79 ± .54) than in the perception task (M = .43 ± .17), $t(34) = -2.71$, $p < .05$. Therefore, as participants were less accurate in the production task than in the perception task, the calibration margin was basically more important in Experiment 2 than in Experiment 1.

3.4 Conclusions

The results of the numerosity production task also replicated patterns of performance previously found in studies on numerical estimation. Firstly, as in Experiment 1, the scalar variability principle well described our data. Secondly, participants overestimated target magnitudes (e.g., Crollen et al., 2011).

Our data also extended Izard and Dehaene's observation (Izard & Dehaene, 2008), according to which symbolic feedback can improve numerosity perception judgments. In this second experiment, we showed that non-symbolic feedback could improve participants' performances in a numerosity production task. However, whereas the feedback effect in Experiment 1 reflected the place-coding feature of the symbolic numerical representation, the feedback effect in Experiment 2 appeared to reflect the summation-coding feature of the non-symbolic numerical representation. Indeed, in the same way as a dot pattern primes all target values that are smaller or equal to the prime value (Roggeman et al., 2007), our results demonstrated that non-symbolic feedback calibrates numerosity production judgments by more when the

target numerosity is smaller than the feedback value than when it is larger than the feedback value.

4. General discussion

Previous research on numerical cognition has highlighted the fact that the symbolic and non-symbolic notations could produce a different pattern of activation on a common magnitude representation: a place-coded pattern on the one hand, and a summation-coded pattern on the other (Roggeman et al., 2007; Verguts & Fias, 2004). The present work showed that numerosity estimation performances could reflect these differential codings. More particularly, it appeared that the format used to present the feedback trial and the participants' responses influenced the shape of the calibration effect: when a symbolic feedback was provided to participants in a numerosity perception task (i.e., symbolic responses), the calibration effect appeared to be of the place-coding type (i.e., inverted V-shaped pattern of calibration). Conversely, when a non-symbolic feedback was provided to participants in the numerosity production task (i.e., non-symbolic responses), the calibration effect appeared to be of the summation-coded type (i.e., step-like pattern of calibration).

These findings extended the results demonstrated by Roggeman et al. (2007) in that they showed, with a different behavioural task, that the representation of large numerical values depends on two processing systems, exactly like the representation of small numerical values. While the summation-coding system has been observed with numerosities ranging from 1 to 5 in a priming paradigm (Roggeman et al., 2007) and in a simulation study (Verguts & Fias, 2004), it has not been observed with canonical configurations of dots (Herrera et Macizo, 2008) and fingers (Di Luca et al., 2010). In this paper, the summation-coding system was found in an experimental

paradigm involving the calibration of the numerosity estimation process and large numerical values ranging from 25 to 95. In order to account for these discrepancies, we will refer to Peirce's theory of signs. For Peirce (1931-1935), the symbol is a sign which is connected to its object thanks to a convention and through an instinct or intellectual act that takes the symbol as representing its object, without the need of any action to occur. While canonical configurations of dots and fingers are close to the meaning that Peirce gave to a symbol, the random configurations of dots used in our study are far from this meaning. The difference between conventional and non-conventional physical numerosity seems therefore to be an important factor in determining the activation of magnitude information: while conventional configurations of numerosity have a symbolic meaning and thus activate a place coding-representation of numbers, non-conventional configurations are not directly associated with a symbolic meaning and thus activate a summation-coding representation of numbers.

The possibility of calibrating participants' estimation performances was previously studied by Minturn and Reese (1951) and Izard and Dehaene (2008). However, in these two studies, the authors only demonstrated that it was possible to calibrate the numerosity perception performances by giving symbolic feedback to the participants. The data of our two experiments extended this result by showing that non-symbolic feedback could also improve numerosity production performances. Until now, however, we do not know the extent of the calibration process: does the calibration on one kind of stimuli (e.g., configurations of dots) also calibrate the numerosity estimation of another kind of stimuli (e.g., sequences of sounds)?

The presence of a larger estimation error in the production task by comparison to the perception task might stem from the precision of each representation: in the

perception task, responses are given following the activation of a linear precise representation, while in the production task, responses are given following the activation of a logarithmic approximate representation. A similar conclusion was reached by Mundy and Gilmore (2009) in a study examining the mapping abilities of 6- and 8-year old children in which children were instructed to map a target quantity to one of two alternative choices. In half the trials, a target Arabic number was presented followed by two sets of dots (symbolic to non-symbolic numerical mapping). In the other half, a target set of dots was presented, followed by two alternative Arabic digits (non-symbolic to symbolic numerical mapping). While children were unable to map a symbol to a non-symbolic quantity when the ratio was .67, they were able to do the mapping in the opposite direction at the same ratio. According to the authors, this asymmetry originates from the level of precision of each representation: whereas the symbolic to non-symbolic version of the mapping task involved one activation of the accurate symbolic representation (for the target Arabic digit) and two activations of the approximate non-symbolic representation (for the two alternative sets of dots), the non-symbolic to symbolic version of the task involved one activation of the approximate representation and two activations of the accurate numerical representation. Children were therefore more accurate in the non-symbolic to symbolic numerical mapping because this condition involves less approximation than the symbolic to non-symbolic numerical version of the task.

To sum up, our study extended three major observations in the field of numerical cognition: **(1)** the processing of large numerical values seems to be supported by two representational systems, exactly like the processing of small numerical values; **(2)** the differences between these two representational systems are reflected, at a behavioural level, in numerosity estimation tasks; **(3)** it is possible to

calibrate not only non-symbolic to symbolic numerical mapping but also symbolic to non-symbolic numerical mapping. A question to be addressed in future research is whether symbolic feedback could improve numerosity production judgments and, inversely, whether non-symbolic feedback could improve numerosity perception judgments. If calibration is found in such cases, it could be particularly interesting to examine how this effect is related to the place and summation-coded representations of numbers.