AUTISM AND INNER SPEECH: A COMPUTATIONAL MODEL OF LANGUAGE FUNCTIONS IN AUTISTIC FLEXIBLE BEHAVIOUR
Supplementary Materials

Giovanni Granato*
Laboratory of Computational Embodied Neuroscience
Institute of Cognitive Sciences and Technologies
National Research Council of Italy,
Rome, Italy
School of Computing, Electronics and Mathematics
University of Plymouth
Plymouth, U.K.
giovanni.granato@istc.cnr.it

Anna M. Borghi
Department of Dynamic and Clinical Psychology
Sapienza University of Rome
Institute of Cognitive Sciences and Technologies
National Research Council of Italy
Rome, Italy
anna.borghi@uniroma1.it

Andrea Mattera
Laboratory of Computational Embodied Neuroscience
Institute of Cognitive Sciences and Technologies
National Research Council of Italy
Rome, Italy
andrea.mattera@istc.cnr.it

Gianluca Baldassarre
Laboratory of Computational Embodied Neuroscience
Institute of Cognitive Sciences and Technologies
National Research Council of Italy
Rome, Italy
gianluca.baldassarre@istc.cnr.it

July 23, 2021

ABSTRACT

Experimental and computational studies propose that inner speech boosts categorisation skills and executive functions, making human behaviour more focused and flexible. In addition, many clinical studies highlight a relationship between poor inner-speech and an executive impairment in autism spectrum condition (ASC), but contrasting findings are reported. Here we investigate the latter issue through a previously implemented and validated computational model of the Wisconsin Cards Sorting Tests. In particular, the model was applied to detect the individual differences in cognitive flexibility and inner speech contribution in ASC and neurotypical participants. Our results suggest that the use of inner-speech increases along the life-span of neurotypical participants but is absent in ASC ones. Although we found more attentional failures in autistic children/teenagers and more perseverative behaviours in autistic young/older adults, only ASC children and ASC older adults exhibited a lower performance than matched control groups. Overall, our results corroborate the idea that the lower use of inner speech in ASC teenagers and young adults is compensated by alternative cognitive strategies (e.g., visual thinking), but it could represent a disadvantage for children (for the missing support of development) and older adults (for the missing compensation of cognitive decline). Moreover, the results suggest that cognitive-behavioural therapies should focus on developing inner speech skills in ASC children as this could provide cognitive support along their whole life span.

1 Computational details of the model

Environment The cards we used are polygons with a unique combination of three visual dimensions (colour, form, and size), each having one of four possible attributes: colour (red, green, blue, yellow); form (square, circle, triangle, bar); size (large, medium-large, medium-small, small). There are thus $4^3 = 64$ combinations (cards) of attributes. We created a simulated environment composed by the objects (cards) which the model can visually explore (visual search) and on which it can executes a physical action (displacement).

Visual sensor The visual sensor returns a $28 \times 28 \times 3$ RGBY pixel matrix, representing a limited portion of the whole virtual table. The visual sensor is actively moved, in a top-down way (visual search), toward the deck and then sequentially toward the target cards. These matrices are then flattened in a vector of 2352 elements and represent the perceptual input to the model.

Working-memory The working-memory is formed by three recurrent units, each having a self-connection, which can acquire a continuous value ranging in $[0, 1]$. The activation of the each unit is characterised by an internal decay toward a baseline (0.5) and is described by the following equation:

$$m_{l,t} = (1 - \phi) \cdot m_{l,t-1} + \phi \cdot \alpha = m_{l,t-1} + \phi(-m_{l,t-1} + \alpha) \tag{1}$$

where $m_{l,t}$ is the value related to a losing unit l ($l \in 1, 2, 3;\ l \neq s$, where s is the selected unit considered below) at time t, $1 - \phi$ is the strength of the recurrent connection, and $\alpha = 0.5$ is the baseline value to which the memory unit activation converges. The activation of each unit represents the likelihood of selection that the system assigns to each of the three possible matching rules of the task related to colour, form, and size. The parameters ϕ is a critical parameter of the model investigated in the simulations.

Motivational component This component is supported by a reinforcement learning algorithm. In particular it receives the external feedback signal (a binary value in $\{0, 1\}$) and subsequently affects the activation of the unit encoding the last selected and used rule, as follows:

$$m_{s,t} = (1 - \mu) \cdot m_{s,t-1} + \mu \cdot r = m_{s,t-1} + \mu(-m_{s,t-1} + r) \tag{2}$$

where $m_{s,t}$ is the new activation of the rule unit, $s \in \{1, 2, 3\}$ is the index of the selected rule, $m_{s,t-1}$ is the current activation of the unit, $(1 - \mu)$ is the strength of the unit recurrent connection, μ regulates the impact of the feedback on the memory, and r is the feedback signal that is equal to 1 in case of positive feedback (correct matching of the deck card and target card) and 0 otherwise. The parameter μ is set to a fixed value of 0.7 for positive feedback and to a variable value for the negative feedback. The latter value is a critical parameter of the model investigated in the simulations.

Hierarchical perceptual component This component is supported by a deep generative model, in particular a *Deep Belief Network* (DBN, [1]) composed of two stacked *Restricted Boltzmann Machines* (RBM). We trained the first RBM, composed of the input layer and the first hidden layer of DBN, with a classical unsupervised learning algorithm for this model (*contrasting divergence*, [2]). We trained the second RBM, composed by the first and second hidden layers of the DBN, with a modified version of the original algorithm that allows us to alter the reconstructions of original inputs to obtain prototypical representations of input image features on which the system focuses on (e.g., in case of a focus on colour, a red triangle given as input is reconstructed as a shapeless red blob). This modification causes the emergence of three groups of units in the last layer of DBN (its second hidden layer), each corresponding to specific visual categories of the input (first four units for colour: red, green, blue, yellow; second four units: square, circle, bar, triangle; third four units: small, medium-small, medium-large, large). The model is able to 'reconstruct' ('generate') the original input trough a bidirectional activation from the input layer, to the hidden layer, and then back to the input layer. In particular, the selector and manipulator considered below are able to select one category (one group of four units), and one attribute within it (one neural unit), to produce the prototypical rule-based reconstruction of images mentioned above.

Selector and manipulator components The selector is supported by a $softmax$ function, a winner-take-all (WTA) function that receives the values from the working memory as input, and chooses the matching rule as follows:

$$Pr(k = s) = \frac{\exp(m_k \,/\, \tau)}{\sum_{q=1}^{3} \exp(m_q \,/\, \tau)} \tag{3}$$

where $Pr(k = s)$ is the probability that the matching rule k ($k \in 1, 2, 3$) is selected ($k = s$). The parameter τ of the $softmax$ function, called 'temperature', regulates the randomness of the selection and is the third important

parameter manipulated in the simulations. A high value of τ causes a high randomness/exploration of the decisions. The probabilities $Pr(\cdot)$, summing up to 1, are used to stochastically select the matching rule to use. The manipulator is composed by two layers of 3 units, linked with one by one negative projections. Each unit of the second layer is always active and has negative projections to a specific group of the last layer of the perceptual component, so the activation of a specific unit in the first layer of the manipulator causes a disinhibition of the corresponding group of the last layer of the perceptual component. Moreover, the manipulator implements a *Hard-max* function leading to select only one unit (attribute) within the each group (category) of four units.

Verbal component This component is supported by a multi-layer perceptron (MLP), formed by 4 input units, 10 sigmoid hidden units, and 3 output linear units. In particular it receives one-to-one connections from the selector units and sends one-to-one connections to the WM units. This process is in particular implemented as follows:

$$m_t = m_{t-1} + \lambda \cdot L_t \tag{4}$$

where m_t is the new activation of a WM rule unit, m_{t-1} is the current activation of the WM unit, λ represents the strengths of the one-to-one connection weights linking the language component output-layer units to the WM units, L_t is the current activation of the language component output layer caused by the previous selector units' activation (this time mismatch implies that the component implements a phonological memory).

The input-layer is formed by 4 units, i.e. the selector winner-takes-all *one-hot vector* activation and the binary incorrect/correct match feedback encoded with respectively 0/1. We trained the MLP to activate the output-layer 3 units as follow: the unit corresponding to the selected rule learned to produce a $-1/+1$ value based on the match/mismatch feedback; the other two units activated with 0. The language component is activated two times to simulate: (a) the phonological-loop working memory; (b) the feedback-dependent verbal update of the main working memory. In the first activation, the component input layer is activated by the one-hot code of the selector while its feedback unit is activated with 1 (meaning 'maintenance of the current rule'). In the second activation, the component input layer is activated by the selector activation, but in this case the feedback unit value is activated on the basis of the external feedback (0/1), obtained after the action execution (displacement of the card). The contribution of language to the working memory is regulated by a coefficient λ that ranges in $[0, 1]$ and represents the strengths of the one-to-one connection weights linking the language component output layer to the main working memory units. The coefficient λ is the fourth and last important parameter regulating the functioning of the model and investigated in the simulations. The language MLP component is trained before the experiments illustrated in the main text with the backpropagation (supervised learning)

Visual comparator This component is supported by a function that computates the Euclidean distance between the two reconstructed images corresponding to the deck card and the currently-foveated target card returned by the perceptual component. It returns a Boolean value representing the result of the comparison ('same'/'not same').

Motor component This component allows (a) the top-down visual search, i.e. the saccades corresponding to the displacement of visuals sensor, and (b) the interaction of the model with environment (displacement of the deck card from the deck to specific target card). The first mechanism receives the position (Cartesian coordinates) of the deck card and the target cards and displaces the visual sensor on them in a sequential manner. The second mechanism receives the position of the deck card and of the matched target card (Cartesian coordinates) and displaces the deck card toward the position of the matched target card.

2 Results

2.1 Fitting results and comparison between the behaviour of the models and of human groups: models validation details

We adopted the same procedure corroborated in [3] to execute the parameters search, aimed to find the parameters of the models that produce the behaviour that best fits those of human populations. In particular we randomly-sampled $3,000$ combinations of parameters each drawn with a uniform distribution in the following ranges: ϕ: $(0.0, 1.0)$; μ: $(0.0, 1.0)$; τ: $(0.0, 0.3)$; λ: $(0.0, 1.0)$. For each parameter combination, we then performed 30 simulations of the task, so obtaining an average value of the WCST indices. We finally computed the Minimum Squared Errors (MSEs) between the WCST indices of the models and human population, as follows:

$$MSE = \frac{||\mathbf{y} - \mathbf{y}'||_2^2}{n} \tag{5}$$

where $\mathbf{y}$ is the vector of mean indices of the human group, $\mathbf{y}'$ is the vector of mean indices of the considered parameter combination, $||\cdot||_2^2$ is the square of the L2 norm, and n is the length of vectors.

Table S1 shows the MSEs for each experimental group while following plots show the comparisons (t-tests) between the human groups and the model groups for each index. Mostly indices are not statistically different, suggesting that the behaviour of the eight models fits the corresponding human populations.

Minimum Squared Errors (MSEs)

	Control	ASC	Means
Children	$1.2 * 10^{-4}$	$6.9 * 10^{-5}$	$9.5 * 10^{-5}$
Teenagers	$4.0 * 10^{-5}$	$2.5 * 10^{-5}$	$3.25 * 10^{-5}$
Young adults	$4.4 * 10^{-5}$	$1.2 * 10^{-4}$	$8.2 * 10^{-5}$
Old adults	$0.70 * 10^{-5}$	$3.5 * 10^{-5}$	$2.1 * 10^{-5}$
Means	$5.3 * 10^{-5}$	$6.2 * 10^{-5}$	$5.8 * 10^{-5}$

Table S1: Minimum Squared Errors (MSEs) of the models that produce the best fit of the data on the WCST indices.

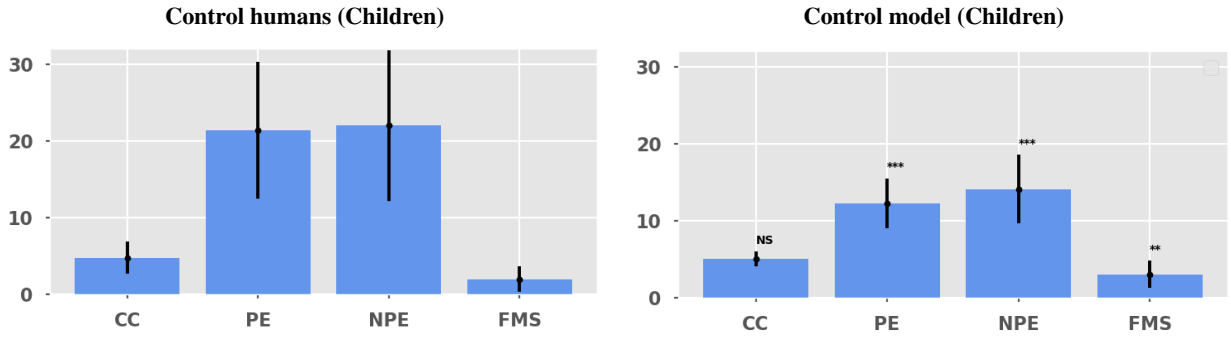

Figure S1: Children (control condition): comparison between the control model group and the control human group of children (** indicates a statistical significance of $p < 0.01$).

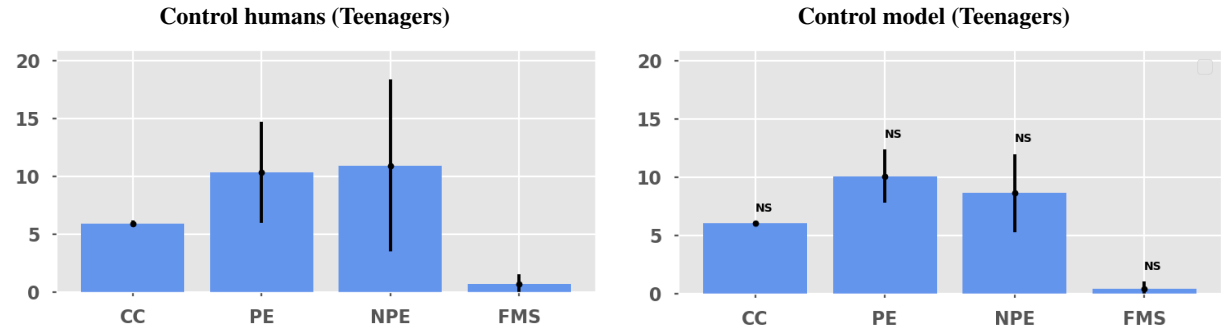

Figure S2: Teenagers (control condition): comparison between the control model group and the control human group of teenagers (** indicates a statistical significance of $p < 0.01$).

2.2 Comparison between the behaviour of different age groups (intra-condition analysis): post-hoc tables

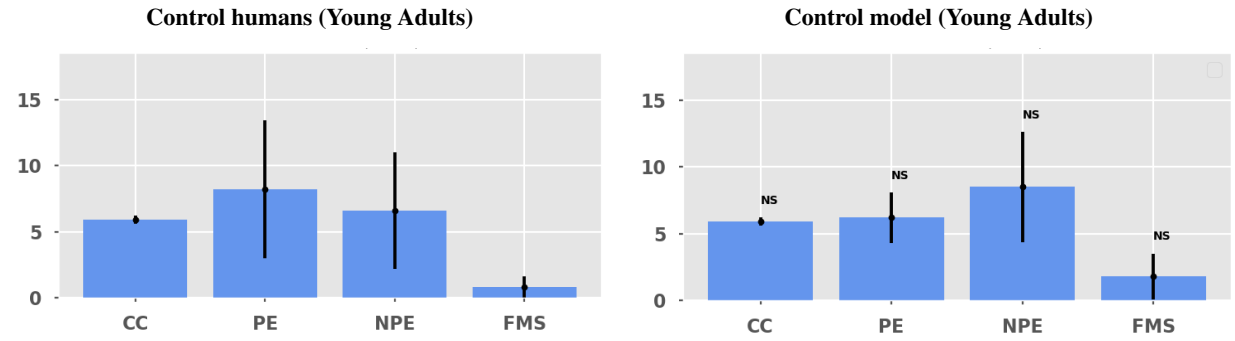

Figure S3: Young Adults (control condition): comparison between the control model group and the control human group of young adults (** indicates a statistical significance of $p < 0.01$).

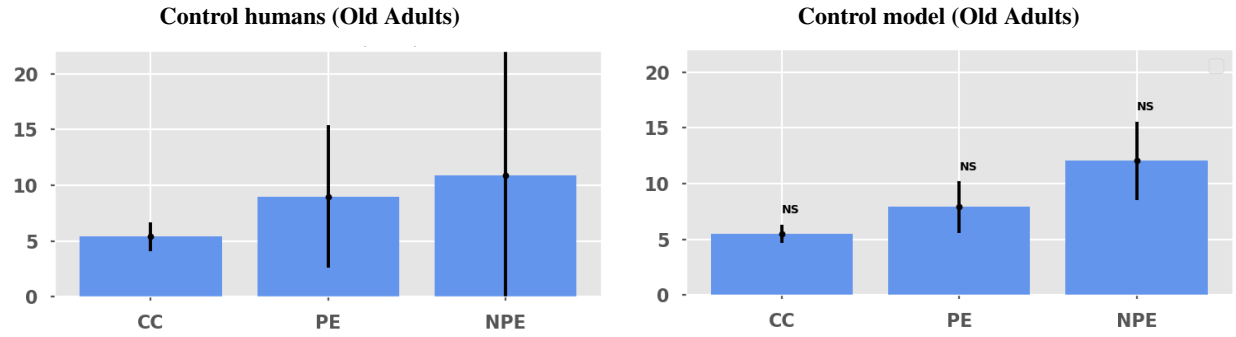

Figure S4: Old adults (control condition): comparison between the control model group and the control human group of old adults (** indicates a statistical significance of $p < 0.01$).

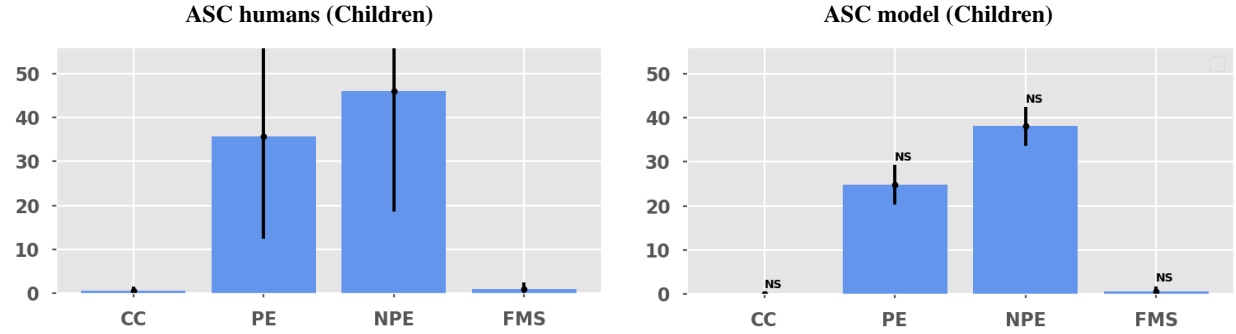

Figure S5: Children (ASC condition): comparison between the autism spectrum condition model group and the Asperger human group of children (** indicates a statistical significance of $p < 0.01$).

Post-hoc tests (CC, control condition)

	Children	**Teenagers**	**Young adults**	**Old adults**
Children	//	$p < 0.01$	$p > 0.05$ (NS)	$p > 0.05$ (NS)
Teenagers	//	//	$p > 0.05$ (NS)	$p > 0.05$ (NS)
Young adults	//	//	//	$p > 0.05$ (NS)
Old adults	//	//	//	//

Table S2: The table shows the post hoc multiple comparisons (t-test with Bonferroni correction) on CC index of control models. NS = not significant.

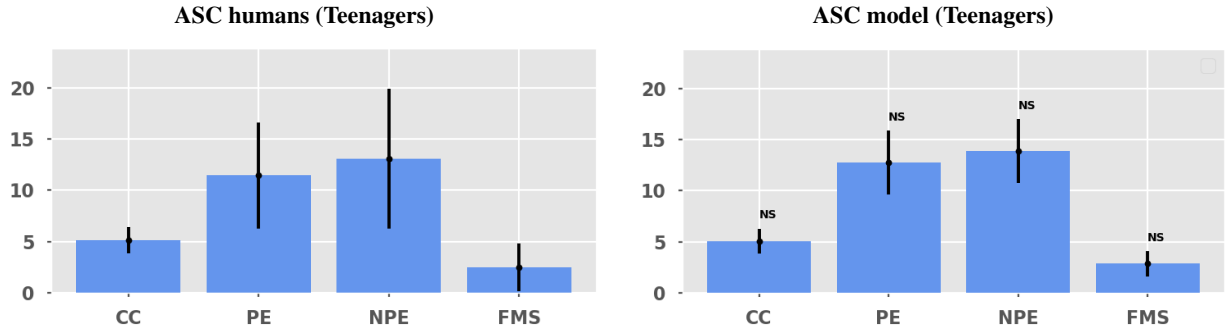

Figure S6: Teenagers (ASC condition): comparison between the autism spectrum condition model group and the a human group of teenagers (** indicates a statistical significance of $p < 0.01$).

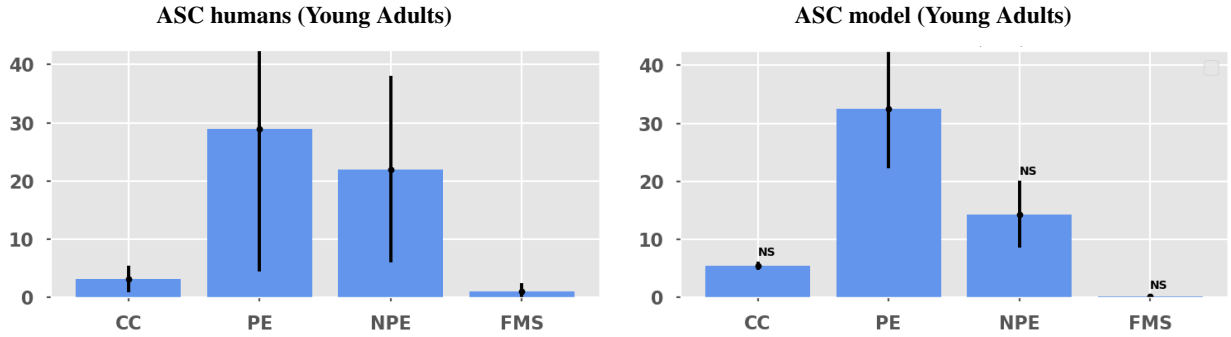

Figure S7: Young Adults (ASC condition): comparison between the autism spectrum condition model group and the a human group of young adults (** indicates a statistical significance of $p < 0.01$).

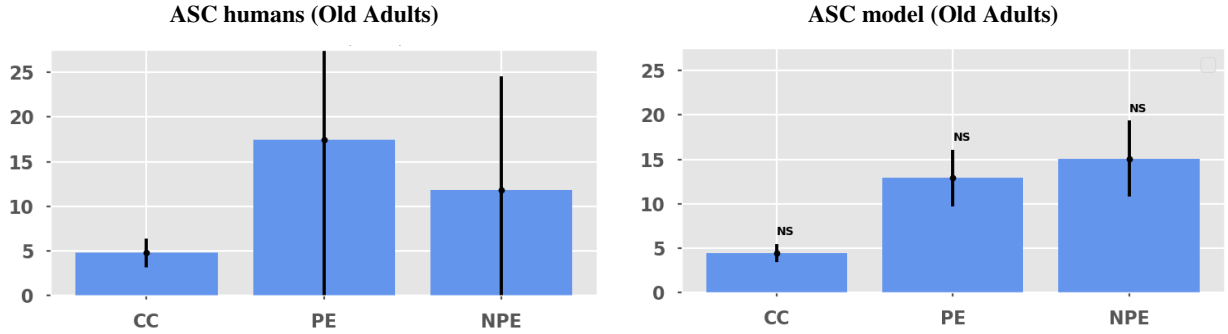

Figure S8: Old Adults (ASC condition): comparison between the autism spectrum condition model group and the a human group of old adults (** indicates a statistical significance of $p < 0.01$).

Post-hoc tests (CC, ASC condition)

	Children	**Teenagers**	**Young adults**	**Old adults**
Children	//	$p < 0.001$	$p < 0.001$	$p < 0.001$
Teenagers	//	//	$p > 0.05$ (NS)	$p > 0.05$ (NS)
Young adults	//	//	//	$p > 0.05$ (NS)
Old adults	//	//	//	//

Table S3: The table shows the post hoc multiple comparisons (t-test with Bonferroni correction) on CC index of ASC models. NS = not significant.

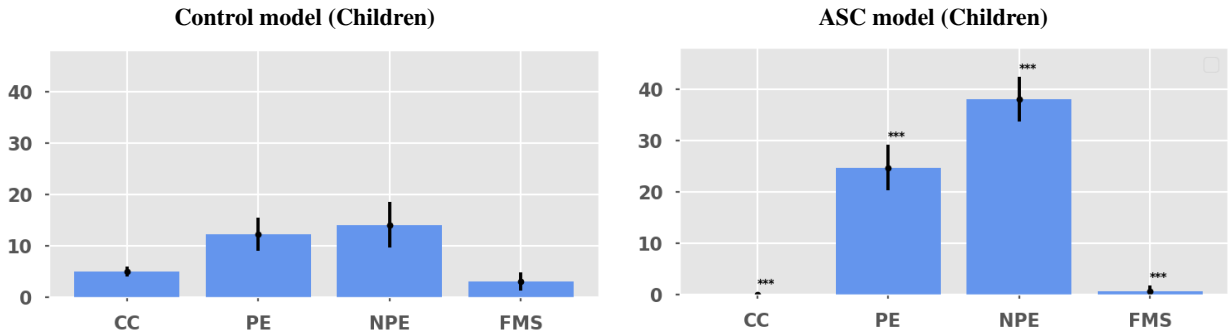

Figure S9: Children (Control-ASC conditions): comparison between the Control model and the autism spectrum condition model of children (** indicates a statistical significance of $p < 0.01$).

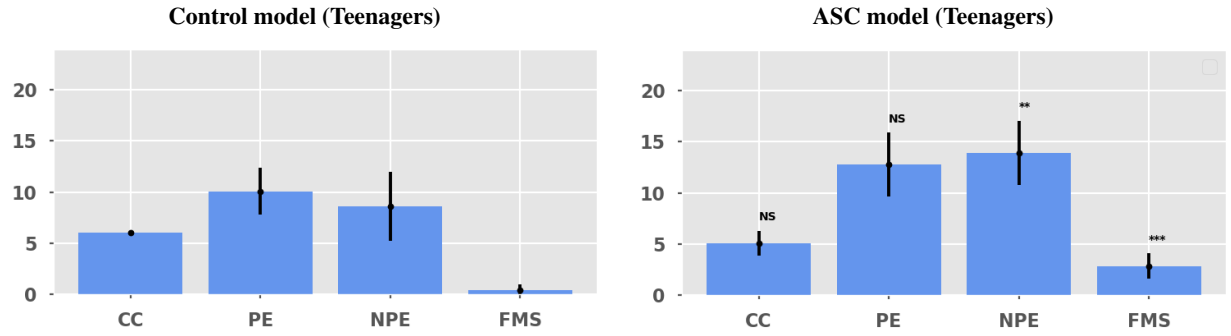

Figure S10: Teenagers (Control-ASC conditions): comparison between the Control model and the autism spectrum condition model of teenagers (** indicates a statistical significance of $p < 0.01$).

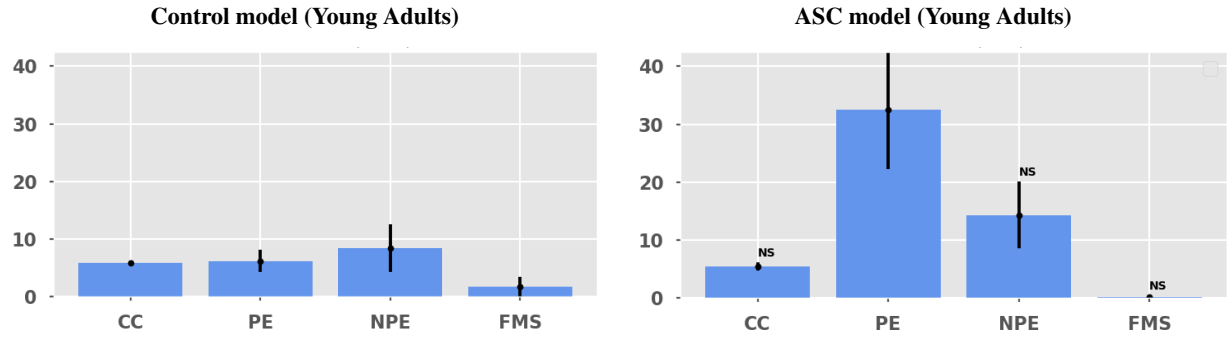

Figure S11: Young adults (Control-ASC conditions): comparison between the Control model and the autism spectrum condition model of young adults (** indicates a statistical significance of $p < 0.01$).

Post-hoc tests (PE, control condition)

	Children	Teenagers	Young adults	Old adults
Children	//	$p > 0.05$ (NS)	$p < 0.001$	$p < 0.001$
Teenagers	//	//	$p > 0.05$ (NS)	$p > 0.05$ (NS)
Young adults	//	//	//	$p > 0.05$ (NS)
Old adults	//	//	//	//

Table S4: The table shows the post hoc multiple comparisons (t-test with Bonferroni correction) on PE index of control models. NS = not significant.

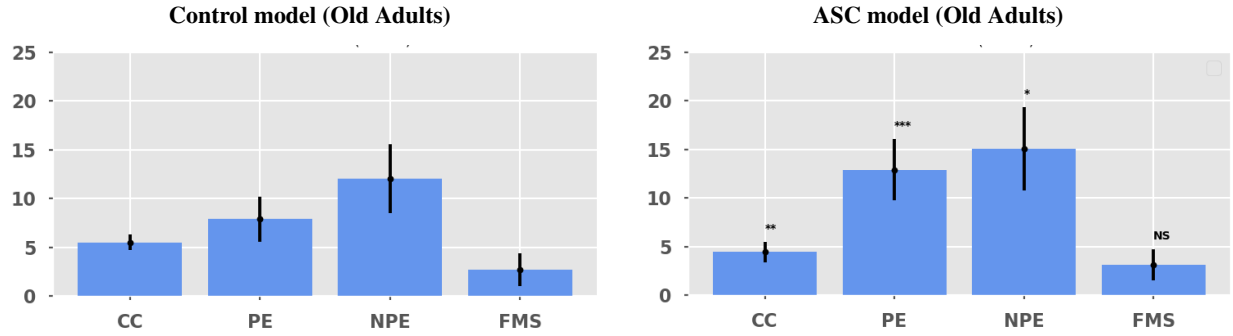

Figure S12: Old adults (Control-ASC conditions): comparison between the Control model and the autism spectrum condition model of old adults (** indicates a statistical significance of $p < 0.01$).

Post-hoc tests (PE, ASC condition)

	Children	Teenagers	Young adults	Old adults
Children	//	$p < 0.001$	$p < 0.001$	$p < 0.001$
Teenagers	//	//	$p < 0.001$	$p > 0.05$ (NS)
Young adults	//	//	//	$p < 0.001$
Old adults	//	//	//	//

Table S5: The table shows the post hoc multiple comparisons (t-test with Bonferroni correction) on PE index of ASC models. NS = not significant.

Post-hoc tests (NPE, control condition)

	Children	Teenagers	Young adults	Old adults
Children	//	$p < 0.001$	$p < 0.01$	$p > 0.05$ (NS)
Teenagers	//	//	$p > 0.05$ (NS)	$p > 0.05$ (NS)
Young adults	//	//	//	$p > 0.05$ (NS)
Old adults	//	//	//	//

Table S6: The table shows the post hoc multiple comparisons (t-test with Bonferroni correction) on NPE index of control models. NS = not significant.

Post-hoc tests (NPE, ASC condition)

	Children	Teenagers	Young adults	Old adults
Children	//	$p < 0.001$	$p < 0.001$	$p < 0.001$
Teenagers	//	//	$p > 0.05$ (NS)	$p > 0.05$ (NS)
Young adults	//	//	//	$p > 0.05$ (NS)
Old adults	//	//	//	//

Table S7: The table shows the post hoc multiple comparisons (t-test with Bonferroni correction) on NPE index of ASC models. NS = not significant.

Post-hoc tests (FMS, control condition)

	Children	**Teenagers**	**Young adults**	**Old adults**
Children	//	$p < 0.001$	$p > 0.05$ (NS)	$p > 0.05$ (NS)
Teenagers	//	//	$p > 0.05$ (NS)	$p < 0.01$
Young adults	//	//	//	$p > 0.05$ (NS)
Old adults	//	//	//	//

Table S8: The table shows the post hoc multiple comparisons (t-test with Bonferroni correction) on FMS index of control models. NS = not significant.

Post-hoc tests (FMS, ASC condition)

	Children	**Teenagers**	**Young adults**	**Old adults**
Children	//	$p < 0.001$	$p > 0.05$ (NS)	$p < 0.001$
Teenagers	//	//	$p < 0.001$	$p > 0.05$ (NS)
Young adults	//	//	//	$p < 0.001$
Old adults	//	//	//	//

Table S9: The table shows the post hoc multiple comparisons (t-test with Bonferroni correction) on FMS index of ASC models. NS = not significant.

References

[1] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural computation. 2006;18(7):1527–1554.

[2] Hinton GE. A practical guide to training restricted Boltzmann machines. In: Neural networks: Tricks of the trade. Springer; 2012. p. 599–619.

[3] Granato G, Borghi AM, Baldassarre G. A computational model of language functions in flexible goal-directed behaviour. Scientific reports. 2020;10(1):1–13.