Motifs and Logic Gates in Small Recurrent Neural Networks

Lindsay Stolting

April 28, 2022

Introduction

One goal at many scales of neuroscience research is to establish a relationship between structure and function. These types of relationships are often well-understood in engineering contexts, where structure is intentionally designed with function in mind. While it is unclear how much we can expect to generalize our way of understanding man-made machines to biology, and vice versa, many argue that the functional building blocks of any computational system must be the same, whether instantiated by a man-made machine or a biological organism (Jonas and Kording, 2017; Marr and Poggio, 1976). Brains and digital computers are arguably the most common objects of this comparison (Zylberberg et al., 2011). In their simplest form, computers are composed of various transistors and circuits which are connected together in such a way that they perform logical operations on their binary inputs and generate appropriate binary outputs. In some sense, the most basic units of functional importance in a computer are these so-called logic gates, which, when linked together in intentional configurations, afford computers their full computational functionality.

Brains, too, have many different levels of organization, one of which is small networks composed of few interconnected neurons. One reason to portray small neural networks as units of functional importance in the nervous system is that strong parallels have been drawn between small neural networks and logic gates. Warren S. McCulloch and Walter Pitts put forth this idea in a 1943 paper, speculating several ways in which units composed of 3-5 neurons could instantiate different logical operations and explaining how these networks could be chained together to perform more complex logical calculations (Mcculloch and Pitts, 1943). If we could locate similarly organized units in the brain, would we be justified in describing their functional role as the brain's logic gates? This also brings to mind the more recently developed technique of network motif analysis, which has found that certain three-node configurations are more common in brains and

other naturally-occurring information processing networks than expected by chance (Milo et al., 2002; Gal et al., 2017). And, in the context of gene regulatory networks, certain frequently occurring motif structures have been successfully related to their functional role (Shen-Orr et al., 2002). Might certain motifs in biological neural networks have functional significance as logic gates, as McCulloch and Pitts described?

Before answering this question, it is important to acknowledge that Pitts' and McCulloch's logical calculus of the nervous system rested on several simplifying assumptions. First, neurons were assumed to send only discrete signals, contradicting modern ideas of firing-rate-dependent information transfer, which is continuously valued. Second, all the proposed networks were feed-forward and cycles were disallowed. This was necessary to make the calculus tractable but, McCulloch and Pitts acknowledged, it is an inaccurate representation of neural connectivity. Thus the question arises; does a link between motif structures and logical operations still exist for continuously-valued recurrent neural networks more closely resembling the brain?

To investigate this potential link, I will first ask which three-neuron motif structures, when realized in a continuously-valued and recurrent neural network model, are capable of performing basic logical operations. Then, I will ask whether these motifs still hold their functional significance when embedded in a larger network, or whether at larger scales the structure-function link becomes muddled.

Model and Methods

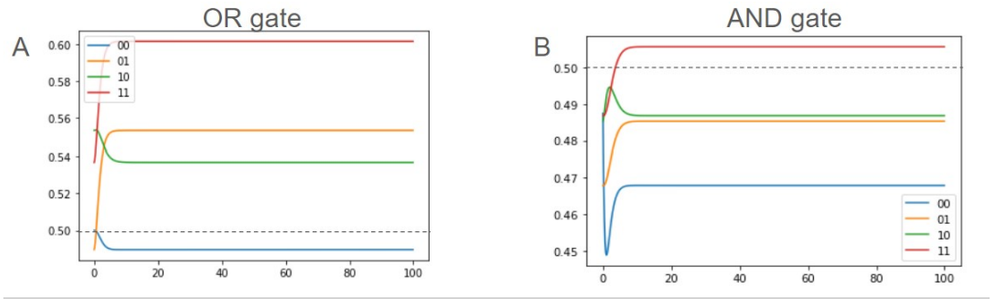
Continuous-time recurrent neural networks (CTRNNs) are a well-suited model to investigate this question because they are, as the name suggests, continuously valued and recurrent. The rate of change of the state of each CTRNN node is governed by the following equation:

$$\tau_i \frac{dy_i}{dt} = -y_i + \sum_{i=1}^{N} w_{ji}\sigma(y_j + \theta_j) + I \tag{1}$$

where y_i represents the state of unit i in the network, τ_i represents its time-constant, θ_i represents its bias, w_{ji} represents the weight connecting unit j to unit i, I represents external input, and $\sigma(x)$ is a sigmoid function, in this case $\sigma(x) = 1/1 + e^{-x}$. For the purposes of this project, all neuron biases will be set to 0 and all time-constants set to 1. Connection weights are allowed to be either excitatory with a value of 1, inhibitory with a value of -1, or absent with a value of 0.

As logical operations, I will focus on the simplest and most widely recognized gate

types: AND and OR. In order to make these operations (which are normally binary and time-independent) realizable in CTRNNs (which are continuously-valued with states that evolve over time), I have operationalized them as follows. Binary inputs are given to two designated input neurons in the network by applying either a tonic input (I) of magnitude 1 (for 1) or no input (for 0). Then, the network is allowed to evolve to its equilibrium point. Because no neuron's self-connection weight can exceed 4, it is not possible for the circuit to be multi-stable (Beer, 1995). After initializing the circuit at $y = 1$ for all neurons and integrating for 1000 time-steps with $dt = 0.1$, equilibrium is assumed to have been reached and the output value ($\sigma(y)$) of a predesignated output neuron is read. If it is above the threshold value of 0.5, this is treated as a "true"/1 output. If it is less than or equal to 0.5, this is treated as a "false"/0 output. To determine if a given circuit is a successful logic gate, this procedure is repeated for every possible combination of inputs (00, 01, 10, 11), and the resulting outputs are compared to the truth table for a desired logic gate (see Figure1C). The circuit is labeled successful only if it gives the correct output for every possible input combination. That is, an OR gate should output 1 only when at least one of the inputs is 1, and an AND gate should output 1 only when both inputs are 1.



INPUT		OUTPUT
A	B	
0	0	0
1	0	1
0	1	1
1	1	1

INPUT		OUTPUT
A	B	
0	0	0
1	0	0
0	1	0
1	1	1

Figure 1: Example time series from the output neuron in (A) a successful OR-gate and (B) a successful AND-gate, correctly classifying each pair of inputs according to (C) the truth table for its gate type

Since each of three nodes has three possible connections, each in three possible states (0, 1, or -1), there are $(3^3)^3 = 19683$ possible motifs. Each one is uniquely defined by an adjacency matrix on the three nodes involved. However, in the case where it is not necessary to distinguish between the three different neurons, many motifs are equivalent

by symmetry. Grouping all the symmetrically equivalent motifs together yields 3411 distinct motif classes (Krauss et al., 2019). The number/index associated with each class roughly corresponds to a progression from every possible connection being inhibitory (-1), through every possible connection being absent (0), to every possible connection being excitatory (1).

3-Neuron Motifs as Gates

First, I tested each of the 19,683 motifs for AND and OR functionality. It was necessary to test the motifs prior to grouping by symmetry because I wanted to distinguish where in the motif the input and output neurons occurred. I identified 52 successful OR gate motifs, but no successful AND gate motifs. The successful OR gate motifs belonged to 25 different classes, with 24 of them being represented twice in the sample due to symmetry across the two input neurons. These class indices were 247, 286, 289, 316, 319, 633, 704, 707, 775, 778, 870, 871, 893, 894, 918, 1373, 1418, 1419, 1437, 1504, 1676, 1679, 1715, and 2053. The other motif class, 1440, was represented 4 times since it also still worked as an OR gate when one of the input neurons was switched with the output neuron (see Figure 2). The sheer number of different OR gate configurations that can be instantiated by a continuous recurrent neural network is already a departure from the previously suggested one-to-one correspondence.

An interesting thing to note about the set of successful OR motifs is that exactly one of the input neurons always excites the output neuron and this constitutes its only incoming connection. This is in contrast to the OR gate configurations proposed by Pitts and McCulloch, in which both input neurons excite the output neuron equally, and no other connections are present. This makes intuitive sense because it should not matter which of the input neurons receives a 1 input; the output neuron should be excited equally in response to either 01 or 10. Implementing Pitts' and McCulloch's simple OR gate configuration on this CTRNN model results in output neuron equilibrium values of $[0.731, 0.774, 0.774, 0.812]$ for inputs of $[00, 01, 10, 11]$, respectively. In other words, while the equilibrium values occur in the correct (ascending) "order" to be an OR gate, they are not split by the chosen threshold value of 0.5.

The motifs that are successful in this case take advantage of another wiring configuration which is made possible because, unlike in the networks proposed by Pitts and McCulloch, the input neurons can connect to each other. Namely, the neuron connected to the output neuron will be excited either by a 1 input to itself or through its connection

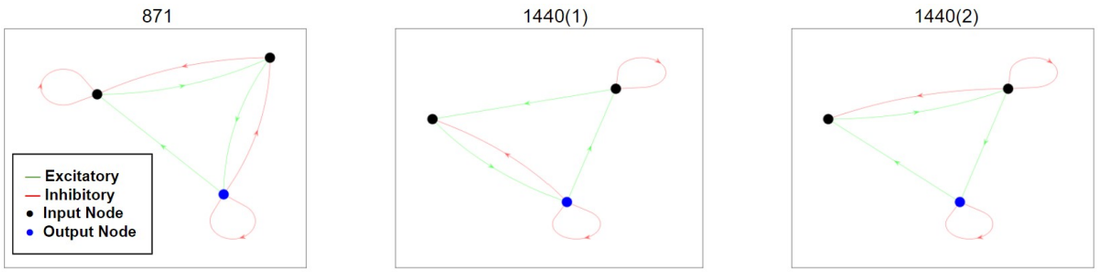


Figure 2: Three examples of successful OR motifs, labeled by motif class index. Notice the output neuron in every successful OR motif has exactly one incoming excitatory connection and it is self-inhibitory.

with the other input neuron, and thus will increase the output neuron's activation level in response to either 01 or 10. Additionally, the output neuron in every successful OR motif was always self-inhibitory. This may play some role in centering the spread of equilibrium values around 0.5, but this hypothesis has yet to be verified. Interestingly, though these two criteria (output neuron with a single excitatory afferent and self-inhibition) are necessary for a successful OR motif, they are not sufficient. Accounting for symmetry across the input neurons, there are 729 motifs that meet this criteria, but most of them are not successful OR gates.

I also performed a 2-sample Kolmogorov–Smirnov (KS) test (which is applicable to probability distributions of discrete random variables) to compare the sample of successful OR motifs to an equally sized random sample from a uniform distribution where all motifs are equally likely to be picked (Noether, 1963). The p-value was 0.0019, indicating that each sample was likely drawn from a different probability distribution, even when a comparison between two random samples from the uniform distribution did not reach significance (p-value 0.475). This suggests that the OR motifs have special attributes and may be more similar to each other than expected by random chance. In addition to the output neuron's single excitatory afferent and inhibitory self-connection, one of these special attributes may be the slightly reduced proportion of excitatory connections. The ratio of excitatory:inhibitory:absent connections in the successful OR motifs was 35:41:41.

Motifs in 5-Neuron Gates

If we claimed to be able to look within the brain's massively recurrent, connected structure and identify configurations of neurons by their logical functionality, it would be necessary to establish that those configurations correspond to that logical functionality even when embedded in the context of other neurons. Therefore, I wanted to investigate whether the same OR motif structures could be identified in a larger network performing

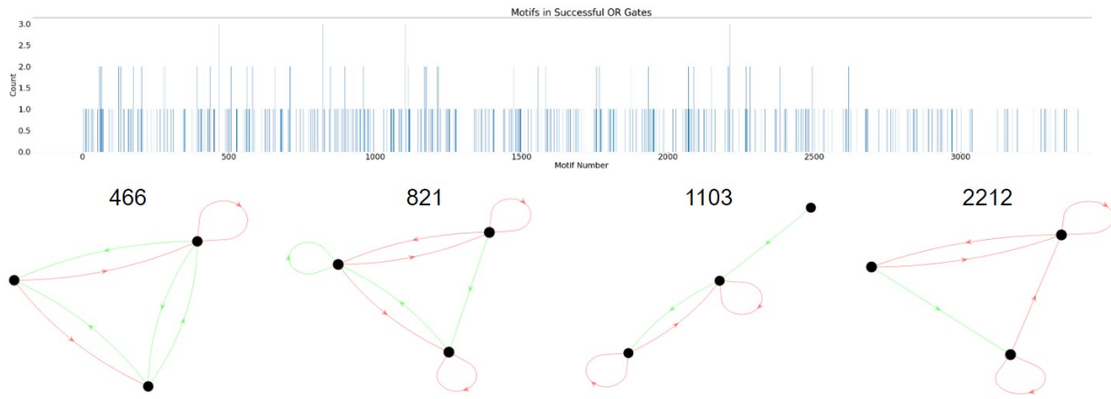


Figure 3: Histogram of represented motif classes in the set of 59 successful 5-neuron OR gates. Graphs of the four most commonly represented classes (3 instances each) are depicted with their indices.

the OR task. To do this, I randomly generated 10,000 5-neuron CTRNNs with an equal probability for each connection of being 0, -1, or 1. Every network had two input neurons, two inter-neurons, and one output neuron. I tested them for OR functionality and found that 59 of them were successful OR gates. Then for these 59 networks, I identified which motif was exhibited by every combination of three neurons in the network and counted how many times each motif appeared in the sample as a whole. I found that 530 different motif classes were present, and no motif appeared more than three times in the sample. The indices of the ones that did so are 466, 821, 1103, and 2212 (see Figure 3). Although I did not keep track of neuron identity (input, inter-neuron, or output) within the motifs (due to combinatorial explosion of possibilities), I noticed that each of these four most popular motifs contained a neuron which received exactly one excitatory input and was self-inhibitory. These were not motifs identified in the first experiment as successful OR gates on their own, but this common feature could be meaningfully related. Of the 25 specific motifs that were identified in the first experiment, only 4 appeared at all in the successful 5-neuron OR gates, and these only appeared once each. From this, it is clear larger networks performing logical operations do not necessarily copy, paste, and embed smaller functional elements. Finally, I compared the distribution of motif classes in the successful 5-neuron OR gates with those in a sample of 59 randomly connected null model networks with the 2-sample KS test. The p-value was $0.0402 < 0.05$, indicating that there may also be something special about the motifs that appear in successful OR gate networks, even though they are not equivalent to the standalone OR motifs. The proportions of absent, excitatory, and inhibitory connections were .320, .314, and .366, respectively.

Next, I generated 10,000 more 5-neuron CTRNNs and tested them for AND function-

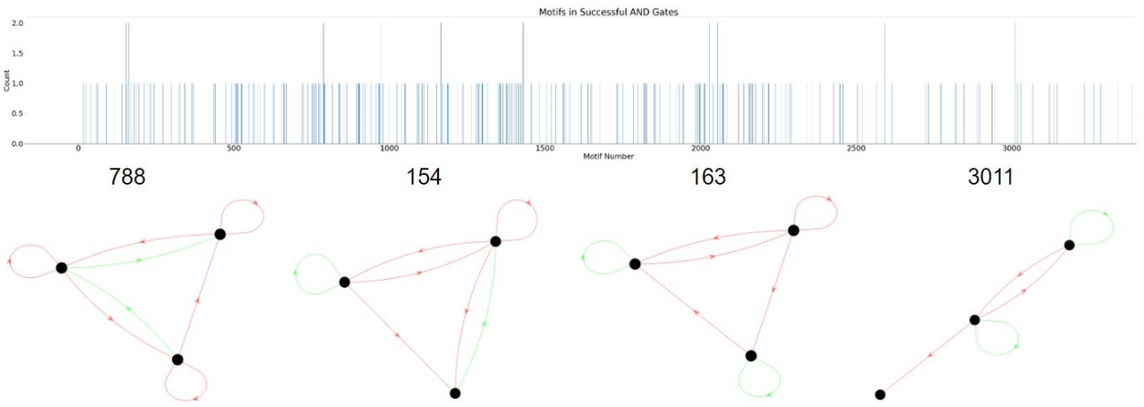


Figure 4: Histogram of represented motif classes in the set of 32 successful AND gates. Graphs of four of the most commonly represented classes (2 instances each) are depicted with their indices.

ality. I found 32 successful AND gates at this larger network size and 308 different motif types were detected. The most common ones occurred twice each and had indices 154, 163, 788, 974, 1166, 1285, 1430, 2028, 2055, 2592, 3011, and 3169 (see Figure 4). Unlike with the OR gates, I do not immediately notice any regularities in this set of motifs. Performing a 2-sample KS test against a sample of 32 randomly connected null graphs yields a p-value of $0.173 > 0.05$. Thus, there is not sufficient evidence to conclude that these two samples were drawn from different distributions, and there may not be anything peculiar about the distribution of motifs present in the 5-neuron AND networks. Directly comparing the AND gate distribution and the OR gate distribution yields a p-value of 0.275, interestingly also indicating no significant difference. Maintaining the trend of a higher proportion of inhibitory connections may partially explain this, as the proportions of absent, excitatory, and inhibitory connections were .3025, .33, and .3675, respectively. Finally, 52 motif types were common across both 5-neuron network samples and, of the 25 motif types that were successful standalone OR gates, 5 appear in the distribution of successful AND gates. This is one more than the number that appear in the distribution of successful OR gates, further challenging the functional significance of these motifs when embedded in a larger network.

Discussion and Future Directions

Although it seems intuitively possible to equate certain three-neuron configurations with logical operations, these intuitions rely heavily on the assumptions of discreteness and seriality. The results of my project suggest that, in networks which are recurrent and have continuously-valued nodes, our intuition fails to so cleanly map structure to function. Results of the first investigation show that a wide variety of structures are capable of

instantiating an OR gate, and that it is impossible to rule-out or rule-in structures based on any one of their features in isolation. Certain topological features may point towards a sub-network's functional role but the correspondence is not perfect and is highly context-dependent.

These results have also brought to light the importance of operational details, both in the context of the model itself and when trying to relate it back to biology. My choice of threshold for the binary output, for instance, likely had a major impact on which motif types were successful. Specifically, the motif types for which the equilibria are centered around 0.5 are more likely to employ a balance of excitatory and inhibitory connections. Broadly, this is what I observed. This requirement of centering is one possible explanation for the universality of self-inhibition in the OR gate output neurons, as well as the relative difficulty of producing an AND gate. On the surface, an AND gate does not seem to be more difficult to produce than an OR gate (both are easily linearly separable functions), except for the fact that an AND gate requires more sub-threshold equilibria, while an OR gate requires more super-threshold equilibria. As a further illustration of the importance of threshold, if I had chosen it to be 0.75 rather than 0.5, McCulloch's and Pitts' own suggestion of an OR motif (which includes only two excitatory connections from each input to output) would have been successful.

Although it is a highly abstract object, the choice of this threshold may have some biological importance. For one thing, neurons are thought to tightly regulate their average firing rate over time and to discourage extremely high or low activation levels (Olypher and Prinz, 2010). Therefore, a threshold value that is too high or too low would be unsustainable, as neurons would be discouraged from crossing it. Additionally, one might imagine that the brain could "chain" logic gates together, such that the output of one gate is an input to another, in order to perform more complex computations. If this were the case, then gate outputs would have to be of an acceptable form and magnitude to serve as inputs. The threshold would have to reliably separate acceptable "true"/1 inputs from acceptable "false"/0 inputs. In reality, of course, this chaining may be hard to achieve given a tendency for diminishing signal strength across synapses, and the heterogeneity of activation thresholds among neurons.

One potentially illuminating future direction would be to attempt chaining, and try to tune this threshold, as well as neural parameters and the magnitude of tonic input applied to the input neurons, so that it is viable. Conversely, one might consider relaxing the requirement of a threshold and selecting instead for circuits which provide outputs

in the correct order of increasing magnitude. For AND and OR gates, this order is $00 < 01, 10 < 11$. In other words, it could be allowed that the threshold be anywhere as long as the set of output equilibria is separable by it in the correct way. By contrast, an XOR gate (which I was unable to produce here) would require the order of the corresponding outputs to be $00, 11 < 01, 10$. Relaxing the requirements in this way may place more of an emphasis on network topology and be less likely to pick up on structural regularities required to conform to the specific threshold (like excitatory/inhibitory balance, self-inhibition, etc.).

Results of the second investigation demonstrate that we must use caution when trying to generalize regularities observed in small network units to a larger scale. The mapping between structure and function is most certainly messy. Although it is tempting to think of networks as modules that can be interconnected without affecting the function of each individual component, this is not always the case. Furthermore, especially in systems like the brain, which can be investigated at so many different scales, a single functional unit can take many shapes, which are not necessarily of the simplest possible form.

Also, this project has helped me to realize the extent to which input and output are imprecise terms when applied to the nervous system. Beyond their defining characteristic of accepting information from outside the system, the "input" neurons also play a major role in computing the output (that is, determining the equilibrium point). The input neurons in perfectly successful logic gates may receive connections from the output neuron, for example. And, if it weren't for this connection, the gate would not function properly. The same is true of output neurons, which may send crucial projections back onto the input neurons or inter-neurons. This is one reason why the computer analogy in neuroscience might be a bit misleading when it comes to questions about information processing. It may be better to think of neural networks as one big dynamical system which gets perturbed in certain ways as a *whole*, and as a consequence the *whole* system changes its state.

Other future directions include keeping track of neuron identity (input, inter-neuron, output) within the motifs of the larger networks. This could reveal regularities in the relationships between these neuron types, even if the relationships between neurons in general were not remarkable. As suggested in class, I could also attempt to evolve networks to exhibit certain gate types, which might be a better way to search the space of possibilities than random sampling. Additionally, having identified some characteristics of a successful OR motif, it could be interesting to check if they appear more often than

expected in the CTRNNs of agents performing computational tasks, or tasks in which an OR comparison is likely to be useful (i.e. size discrimination, catching/avoidance). Lastly, efforts have been made in the field of artificial intelligence to formulate a continuously valued logic, which might be more amenable to implementation in CTRNNs while still being rigorously defined (Preparata and Yeh, 1972).

Overall, through this project I have highlighted the nuance of structure-function relationships in small recurrent neural networks. While it may not be such a simple task to deduce a network's logical function from its topology, there is at least some indication that certain topological features hold specific functional significance. It remains to be seen whether these observed regularities facilitate the logical function of the network, OR other aspects of the task as I have defined it here. However, it is clear that network analysis is a valuable tool for studying computation in a system which makes such adept use of its dynamics AND interconnectedness.

References

Beer, R. D. (1995). On the Dynamics of Small Continuous-Time Recurrent Neural Networks. *Adaptive Behavior*, 3(4):469–509. ZSCC: 0000534 Publisher: SAGE Publications Ltd STM.

Gal, E., London, M., Globerson, A., Ramaswamy, S., Reimann, M. W., Muller, E., Markram, H., and Segev, I. (2017). Rich cell-type-specific network topology in neocortical microcircuitry. *Nature Neuroscience*, 20(7):1004–1013. Number: 7 Publisher: Nature Publishing Group.

Jonas, E. and Kording, K. P. (2017). Could a Neuroscientist Understand a Microprocessor? *PLOS Computational Biology*, 13(1):e1005268. Publisher: Public Library of Science.

Krauss, P., Zankl, A., Schilling, A., Schulze, H., and Metzner, C. (2019). Analysis of structure and dynamics in three-neuron motifs. *Frontiers in Computational Neuroscience*, 13:5. Publisher: Frontiers.

Marr, D. and Poggio, T. (1976). From Understanding Computation to Understanding Neural Circuitry. Accepted: 2004-10-01T20:36:50Z.

Mcculloch, W. S. and Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5:115–133.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827. Publisher: American Association for the Advancement of Science.

Noether, G. E. (1963). Note on the kolmogorov statistic in the discrete case. *Metrika*, 7(1):115–116.

Olypher, A. V. and Prinz, A. A. (2010). Geometry and dynamics of activity-dependent homeostatic regulation in neurons. *Journal of Computational Neuroscience*, 28(3):361–374. ZSCC: 0000036.

Preparata, F. P. and Yeh, R. T. (1972). Continuously valued logic. *Journal of Computer and System Sciences*, 6(5):397–418.

Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics*, 31(1):64–68. Number: 1 Publisher: Nature Publishing Group.

Zylberberg, A., Dehaene, S., Roelfsema, P. R., and Sigman, M. (2011). The human Turing machine: a neural framework for mental programs. *Trends in Cognitive Sciences*, 15(7):293–300.