The Information Bottleneck Principle in Corporate Hierarchies

Cameron Gordon
Australian Institute for Machine Learning
University of Adelaide
cameron.gordon@adelaide.edu.au

Abstract

The hierarchical nature of corporate information processing is a topic of great interest in economic and management literature. Firms are characterised by a need to make complex decisions, often aggregating partial and uncertain information, which greatly exceeds the attention capacity of constituent individuals. However, the efficient transmission of these signals is still not fully understood. Recently, the information bottleneck principle has emerged as a powerful tool for understanding the transmission of relevant information through intermediate levels in a hierarchical structure. In this paper we note that the information bottleneck principle may similarly be applied directly to corporate hierarchies. In doing so we provide a bridge between organisation theory and that of rapidly expanding work in deep neural networks (DNNs), including the use of skip connections as a means of more efficient transmission of information in hierarchical organisations.

1 Introduction

The corporate structure of firms has been the subject of economic enquiry since Coase (1937) [1; 2]. In particular, the structures required for information processing and communication within firms has received much interest from economists, organisational theorists, institutional researchers, cyberneticists, and sociologists [3; 4; 5; 6; 7; 8; 9; 10]. Firms must acquire, process, and act upon information in order to effect a goal or strategic commercial objectives, often under conditions of ambiguity, partial observability, and uncertainty [3; 2]. This requires the coordination of many individuals whose bounded information capacity is substantially lower than that of an organisation as a whole [3; 11; 4; 12]. Employees are typically arranged in a hierarchy with lower levels reporting to higher and so on up until the board [3; 13; 14]. While decision-making and delegation may occur at all levels of an organisation, a formal decision-maker (e.g. a board member or executive) may be separated by several layers of hierarchy from the original source of information relevant to their decision [12; 3]. Indeed, as Simon (1997) notes "information most important to top managers comes mainly from external sources" [3]. Market statistics, supply-chain information, changes to government policy, media reports, customer interactions, and sales figures feature among the external information relevant to decision-making. This aggregated set of information in most cases far exceeds the attention limit of any one individual to process; hence, the question of how to efficiently process, transmit, and filter decision-relevant information is of critical importance for firms [12; 3; 11].

The transfer of relevant information within a firm may be seen principally as a question of signal transmission. The well-developed theory of the *information bottleneck principle* provides a means of analysing the communication of *relevant information* that exists between two signals potentially involving multiple stages of processing [15]. Furthermore, this theory has recently emerged as a powerful tool for understanding the information processing that occurs in the similarly hierarchical structure of deep neural networks (DNNs) [16; 17]. While the hierarchical similarity between DNNs

Information-Theoretic Principles in Cognitive Systems Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

and corporate hierarchies has been noted sporadically within the literature (e.g. [18]), the information theoretic and organisational design consequences of this connection have received little examination.

The primary contribution of this paper is to provide a bridge between organisational theory and rapidly expanding work in deep neural networks and information bottleneck theory. We additionally explore a common component of deep neural networks (skip connections) and consequences for efficient corporate information processing. We view Helbing et al. [12] as the prior literature most related to our own. The authors investigate information flow in hierarchical networks and highlight the importance of side-channels and shortcut connections for efficient information transmission, however they do not extend this to the information bottleneck principle or neural architectures.

2 The Information Bottleneck Principle

The information bottleneck principle introduced in [15] describes conditions for a compressed intermediate representation between two random variable signals (an input and an output), by reference to the mutual information shared between these signals and the representation [16; 15]. The principle states that a compressed intermediate representation should contain the *minimum* information with respect to the input while remaining predictive of the output [15]. Consider two random variables X and Y. The entropy of a random variable X is a measure of its uncertainty [19]:

$$H(X) = -\sum_{x \in X} p(x) \log_2(p(x)). \tag{1}$$

The mutual information between two random variables $I(X;Y)$ symmetrically describes the decrease in the uncertainty in X that occurs given that Y is known (and vice-versa) [19]:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{2}$$

Take an intermediate representation $\hat{X}$ between an input X and output Y. Then $I(X;\hat{X})$ describes the information that is preserved between the input and the representation; and $I(Y;\hat{X})$ that between representation and the output. Using a control parameter β, the information bottleneck principle states that the compressed representation should minimise the Lagrangian [16; 15; 17]:

$$\mathcal{L}[p(\hat{x}|x)] = I(X;\hat{X}) - \beta I(Y;\hat{X}), \tag{3}$$

where $p(\hat{x}|x)$ is a mapping for all $x \in X$ satisfying the probability constraint $\sum_{\hat{x} \in \hat{X}} p(\hat{x}) = 1$. The parameter β controls the rate of compression and may be viewed as a constraint on transmission.

The information bottleneck principle has been applied widely to problems of information processing in decision science [20], psychology [21], and prediction [22]. Importantly, it has been used to analyse information flows in deep feedforward neural networks [16; 17; 23; 22]. A deep feedforward neural network involves a hierarchical processing of layers in which the output of each layer feeds directly to the next. Typically this is written as a composition of functions $f(x) = f_k(f_{k-1}(f \ldots (f_1(x))))$ where x is an input, $f_i(x)$ is a function applied at layer i, and k is the number of layers [24].

Tishby and Zaslavsky [16] describe two properties that enable the information bottleneck principle to be applied to DNNs. The first is the data processing inequality, a fundamental result in information theory [19]. Briefly stated, for a Markov chain $X \to Y \to Z$ it must be that $I(X;Y) \geq I(X;Z)$. That is, that mutual information is non-increasing through processing. This may be seen visually by observing the sets of information shared between intermediate representations (see Figure 1). Within a DNN information flow can be represented as a Markov chain $X \to T_1 \to T_2 \to \ldots T_k \to Y$ where T_i represents the ith hidden layer. The bottleneck representation implies that at convergence each layer will be the minimal representation required to predict the next layer subject to network capacity. Successive neural network layers hence reduce the mutual information shared between the input and output. The second property (invariance of mutual information to invertible transformations) enables the principle to hold for different re-parameterizations and intermediate structural forms [21].

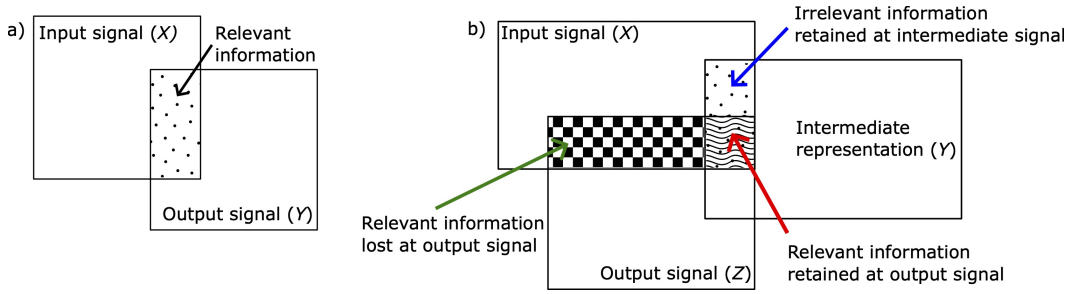

Figure 1: a) The relevant information is the minimal set of information in an input signal that is predictive of an output signal; without processing this corresponds to the mutual information $I(X;Y)$. b) Visualizing the data processing inequality. Taking an intermediate representation forms the Markov chain $X \to Y \to Z$. The mutual information between input signal X and output signal Z must be less than or equal to that between the input signal and the intermediate representation Y.

3 Hierarchies in Corporations

Corporate organisations are typically arranged in a hierarchy, with reporting relationships from subordinates to superiors up to the board [3; 1; 25; 6; 2; 26]. Consider the following simplified model of a corporation with N levels of hierarchy. For illustration, we first consider the simple case in which hierarchies are strict, and information flows only from one hierarchical layer to the next (i.e. information enters the corporation at layer 1 and is passed following processing to layer 2). At each layer L_i some processing occurs (e.g. briefing notes, statistical analysis) with active decision taken by an individual on what information to transmit and what to filter from their superiors. We can note that we have the Markov chain $L_1 \to L_2 \to \cdots \to L_N$, with the data processing inequality $I(L_1;L_2) \geq I(L_1;L_3) \geq \cdots \geq I(L_1;L_N)$. Hence, this structure obeys a data processing inequality. By making the assumption that transformations within layers are invertible (i.e. prior to transmission the input information to that layer may be fully recovered), we can begin to apply the information bottleneck principle to this corporate hierarchy.

A chief information constraint within firm management is the limited attention capacity of managers, who each operate within the constraints of bounded rationality [3; 11; 12]. Hence we may frame an information processing objective at each level in a hierarchy as transmitting the maximally relevant information subject to it being less than the attention capacity of managers at the next level. We can note this is equivalent to filtering or removing irrelevant information from being transmitted. Assume for the purposes of a decision that the attention constraint for a *relevant decision-maker* at level k is α_k. With firm external information X we have following optimisation at level $k-1$ of the organisation to produce a constrained information representation L_k based on the preceding level:

$$\underset{L_{k-1}}{\text{argmin}}\ I(L_{k-2};L_{k-1}) - \alpha_k I(L_{k-1};L_k), \quad (4)$$

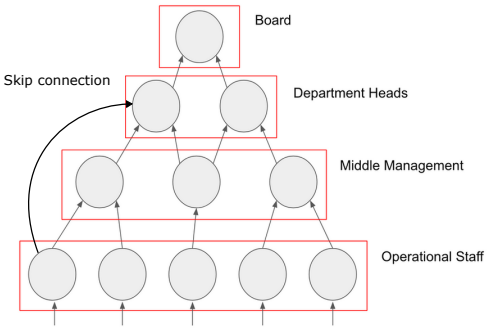

Figure 2: Information flow through a hierarchical organisation. Information entering an organisation at the operational level may have relevance to the board and is transmitted primarily through reporting lines. Skip connections within the organisation (informal and formal) enable lower level representations to be re-transmitted higher in the hierarchy, reducing the risk of incorrect filtering.

which an be recognised as an information bottleneck. Similar analysis can be extended to lower levels in the hierarchy where the attention constraint for the decision-maker is replaced by the sum of attentions of individuals j in the following layer:

$$\underset{L_{k-2}}{\text{argmin}}\ I(L_{k-3}; L_{k-2}) - \sum_j \alpha_{j,k-1} I(L_{k-2}; L_{k-1}). \tag{5}$$

And finally to the operational level where the signal X enters the organisation:

$$\underset{L_1}{\text{argmin}}\ I(X; L_1) - \sum_j \alpha_{j,1} I(L_2; L_1). \tag{6}$$

We can view the objective at each level as *decoding* the information received and *encoding* a compressed representation that can be transmitted through relevant reporting relationships [16; 17].

Aligning the encoding-decoding process at each layer is required to reduce error in prediction and decision-making. This requires the use of top-down *feedback*, a component known to be a critical mechanism for learning across domains [4; 27]. In deep neural networks feedback usually occurs through backpropagation of an error signal [28; 24]. In corporate structures feedback occurs through both formal and informal channels, including performance reviews, incentives, supervision, hiring decisions, and formal direction in order to align individual incentives to strategic goals. [3; 1; 7; 2].

4 Skip Connections in Corporate Hierarchies

Much of the information relevant to corporate objectives is observed at the operational level and must be transmitted up the chain [3]. In passing through several intermediate layers of a strict hierarchy decision-relevant information may be lost due to the data processing inequality, for example through misinterpretation or incorrect filtering by a manager. As Simon [3] writes, "A major communications problem, then, is that much of the information relevant to the decisions at this level originates at lower levels, and may not ever reach the higher levels unless the executive is extraordinarily alert."

One resolution of this issue is that communication networks in corporations are typically *not* strict hierarchies, but instead combine a mixture of formal and informal channels connecting multiple levels in the organisation [3; 12]. It is well-known that different managers selectively attend to different details, and consultation with multiple groups occurs during making business decisions in practice [29; 30; 3]. The presence of shortcut, side-channel, and skip-connections has the effect of combining multiple filtered sources of an incoming signal, enabling error correction and multiple levels of compression to be viewed simultaneously [12]. An example of this is the transmission of a full report alongside a highly compressed executive summary. If relevant information has been incorrectly filtered by the summary it can be recovered through the full report. Similarly, lower-level analysts are often made present while their superiors are reporting to executives. The executive may rely primarily on the compressed representation presented by the superior as a matter of efficiency; and seek additional details, confirmation, or clarification from the subordinate only if required.

The concatenation of a compressed representation with that of an earlier stage in a hierarchy can be recognised in deep learning as a form of skip connection, which have wide use in *ResNets*, ResNeXTs, DenseNets, and related architectures, enabling extremely deep networks to be trained without the loss of signal and vanishing gradient known to affect deep strictly sequential networks [31; 32; 33; 34; 35; 36; 24]. Skip connections can also be viewed as *bridges* which reduce path depth when viewed through the lens of network theory or communication topology [37; 38; 39].

5 Conclusion

Corporations need to filter, compress, and transmit information internally in order effect decisions. In this paper we have argued that the information bottleneck principle may be used to analyse information processing within these hierarchical systems. In doing so we provide a bridge between work in organisational theory and recent advances in the study of neural networks. To an extent these systems face an analogous problem of making decisions under constrained resources: computational and financial [1; 3; 2; 26]. While we have pointed to the example of skip connections as an example of a topological structure in both systems it is an interesting open question whether other features known to improve neural network performance may play an analogous role in organisational design.

Acknowledgements

The author wishes to thank Dr. Lachlan MacDonald, Dr. Hemanth Saratchandran, Dr. Patrick O'Callaghan, and Mr. Keaton Jenner for their helpful suggestions and comments on draft forms of this paper, and Dr. Nan Ye for his help in understanding the details of information bottleneck theory.

References

[1] R. H. Coase, "The Nature of the Firm," *Economica*, vol. 4, no. 16, pp. 386–405, 1937.

[2] H. B. Malmgren, "Information, Expectations and the Theory of the Firm," *The Quarterly Journal of Economics*, vol. 75, no. 3, pp. 399–421, 1961. Publisher: Oxford University Press.

[3] H. A. Simon, *Administrative Behavior, 4th Edition*. Free Press, 1997.

[4] N. Wiener, *Cybernetics: or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press, 2 ed., 1948.

[5] S. Beer, "Brain of the Firm: The Managerial Cybernetics of Organization," 1972.

[6] T. Knapp, "Hierarchies and Control: A New Interpretation and Reevaluation of Oliver Williamson's "Markets and Hierarchies" Story," *The Sociological Quarterly*, vol. 30, no. 3, pp. 425–440, 1989.

[7] O. E. Williamson, "Markets and Hierarchies: Some Elementary Considerations," *The American Economic Review*, vol. 63, no. 2, pp. 316–325, 1973. Publisher: American Economic Association.

[8] B. Markey-Towler, "I, Roboticus Oeconomicus The philosophy of mind in economics, and why it matters," *Cambridge Journal of Economics*, vol. 41, pp. 203–237, Jan. 2017.

[9] B. Markey-Towler, "Rules, perception and emotion: When do institutions determine behaviour?," *Journal of Institutional Economics*, vol. 15, pp. 381–396, June 2019. Publisher: Cambridge University Press.

[10] C. Durugbo, A. Tiwari, and J. R. Alcock, "Modelling information flow for organisations: A review of approaches and future challenges," *International Journal of Information Management*, vol. 33, pp. 597–610, June 2013.

[11] G. Gigerenzer and R. Selten, *Bounded Rationality: The adaptive toolbox*. Cambridge: The MIT Press, 2001.

[12] D. Helbing, H. Ammoser, and C. Kühnert, "Information flows in hierarchical networks and the capability of organizations to successfully respond to failures, crises, and disasters," *Physica A: Statistical Mechanics and its Applications*, vol. 363, pp. 141–150, Apr. 2006.

[13] H. A. Simon, "Designing organizations for an information rich world," in *Computers, communications, and the public interest* (M. Greenberger, ed.), pp. 37–72, 1971.

[14] H. A. Simon, "The Architecture of Complexity," *Proceedings of the American Philosophical Society*, vol. 106, no. 6, pp. 467–482, 1962. Publisher: American Philosophical Society.

[15] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. of the 37-th Annual Allerton Conference on Communication, Control and Computing*, pp. 368–377, 1999.

[16] N. Tishby and N. Zaslavsky, "Deep learning and the information bottleneck principle," in *2015 IEEE Information Theory Workshop (ITW)*, pp. 1–5, Apr. 2015.

[17] R. Shwartz-Ziv and N. Tishby, "Opening the Black Box of Deep Neural Networks via Information," 2017.

[18] S.-I. Marinescu, "Research on Application of Neural Networks in Organizational Management," *Acta Universitatis Cibiniensis. Technical Series*, vol. 65, no. 1, pp. 64–68, 2015.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.

[20] N. Tishby and D. Polani, "Information Theory of Decisions and Actions," in *Perception-Action Cycle: Models, Architectures, and Hardware* (V. Cutsuridis, A. Hussain, and J. G. Taylor, eds.), Springer Series in Cognitive and Neural Systems, pp. 601–636, New York, NY: Springer, 2011.

[21] N. Zaslavsky, C. Kemp, T. Regier, and N. Tishby, "Efficient human-like semantic representations via the Information Bottleneck principle," Tech. Rep. arXiv:1808.03353, arXiv, Aug. 2018.

[22] Z. Goldfeld and Y. Polyanskiy, "The Information Bottleneck Problem and Its Applications in Machine Learning," Tech. Rep. arXiv:2004.14941, arXiv, May 2020.

[23] A. M. Saxe, Y. Bansal, J. Dapello, M. Advani, A. Kolchinsky, B. D. Tracey, and D. D. Cox, "On the information bottleneck theory of deep learning," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, p. 124020, Dec. 2019. Publisher: IOP Publishing.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[25] W. Dessein and R. Holden, "Organizations with Power-Hungry Agents," *The Journal of Law and Economics*, vol. 65, pp. S263–S291, Feb. 2022. Publisher: The University of Chicago Press.

[26] N. J. Foss, "Harald B. Malmgren's Analysis of the Firm: lessons for modern theorists?," *Review of Political Economy*, vol. 8, pp. 349–366, Oct. 1996.

[27] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*. Adaptive computation and machine learning series, Cambridge, Massachusetts: The MIT Press, second edition ed., 2018.

[28] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.

[29] R. M. Cyert, H. A. Simon, and D. B. Trow, "Observation of a Business Decision," *The Journal of Business*, vol. 29, no. 4, pp. 237–248, 1956. Publisher: University of Chicago Press.

[30] D. C. Dearborn and H. A. Simon, "Selective Perception: A Note on the Departmental Identifications of Executives," *Sociometry*, vol. 21, no. 2, pp. 140–144, 1958.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, June 2016. ISSN: 1063-6919.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), Lecture Notes in Computer Science, (Cham), pp. 630–645, Springer International Publishing, 2016.

[33] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987–5995, July 2017.

[34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

[35] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, vol. 28 of *JMLR Workshop and Conference Proceedings*, pp. 1310–1318, 2013.

[36] S. Hochreiter, "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 06, pp. 107–116, Apr. 1998. Publisher: World Scientific Publishing Co.

[37] I. Momennejad, A. Duker, and A. Coman, "Bridge ties bind collective memories," *Nature Communications*, vol. 10, p. 1578, Apr. 2019.

[38] I. Momennejad, "Collective minds: social network topology shapes collective cognition," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 377, p. 20200315, Jan. 2022.

[39] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.