A Theory of Foundational Meaning Generation in Autonomous Systems, Natural and Artificial

Kristinn R. Thórisson[1,2] and Gregorio Talevi[1,3(✉)]

[1] Center for Analysis and Design of Intelligent Agents, Reykjavik University, Reykjavik, Iceland
thorisson@ru.is
[2] Icelandic Institute for Intelligent Machines, Reykjavik, Iceland
[3] School of Science and Technology, University of Camerino, Camerino, Italy
talevigregorio@gmail.com
http://cadia.ru.is

Abstract. The concept of 'meaning' has long been a subject of philosophy and people use the term regularly. Theories of meaning detailed enough to serve as blueprints in the design of intelligent artificial systems have however been few. Here we present a theory of foundational meaning creation – the phenomenon proper – sufficiently broad to apply to natural agents yet concrete enough to be implemented in a running artificial system. The theory states that meaning generation is a *process* bound in the present *now*, resting on the concept of reliable causal models. By unifying goals, predictions, plans, situations and knowledge, it explains how ampliative reasoning and explicit representations of causal relations participate in the meaning generation process. According to the theory, meaning and autonomy are two sides of the same coin: Meaning generation without autonomy is meaningless; autonomy without meaning is impossible.

Keywords: Meaning · Autonomy · Knowledge · Information · Generality · General Machine Intelligence

1 Introduction

The question of why anything can or should 'mean' anything to anyone seems to belong in the set of fundamental questions that scientific and philosophical pursuits should aim to provide an adequate answer to. Surprisingly, in spite of thousands of years of analytical philosophy and close to 200 years of psychological research on this concept, no sufficiently detailed and prescriptive theory exists that can help us build machines that think. Our view is that artificial intelligence, or at the very least its sub-field of general machine intelligence, calls for this question to be addressed.

Philosophy uses the term *foundational meaning* to refer to the *basic* core concept of meaning [7], as in e.g., "the guitar that your father gave you has

K. R. Thórisson et al. (Eds.): AGI 2024, LNAI 14951, pp. 188–198, 2024.
https://doi.org/10.1007/978-3-031-65572-2_21

'a lot of meaning' to you," and when a loved one says you "mean everything" to them, as well as more basic experiences like breaking out in a cold sweat when being robbed at gunpoint: The *meaning* of such situations *to you* – if you are the one having the experience – is *'foundational.'* This use of the term is different from its use when talking about symbols, signs, and language, which has been called 'semantic' meaning. If a parrot were to squawk at you "I'm gonna *MESS* you up!!", our guess is that you would not be frightened. This is because while you generate semantic meaning for the vocal phrase – which happens to constitute a threat – to you the threat itself has no foundational meaning. The only foundational meaning you generate is of 'a parrot making an empty threat.' A parrot capable of messing up a human is a rather frightening thought; if the situation seems funny, it is due to the sharp contrast between the semantic and the foundational meanings. Generating semantic meaning of e.g. text is a classification task; generating foundational meaning involves much more.

Here we present a computational theory of foundational meaning. The theory already has a rudimentary implementation in a cognitive architecture [9,13].[1] In this paper we describe the theory and detail its foundation and formalisms.

2 Related Work

The study of "meaning" has been a central topic in various disciplines, including philosophy, linguistics, and cognitive sciences. Philosophic study of meaning goes back thousands of years: disquisitions and formulations on the nature of the meaning of words, symbols and ideas can already be found in the works of ancient Greek philosophers such as Socrates, Plato and Aristotle [12].

We agree with philosopher David Lewis [7] that it is important to distinguish Foundational and Semantic theories of meaning. The former refers to why anything should 'mean' anything to someone, while the latter refers to translation of some symbol structure into a pragmatic knowledge structure (e.g. the message "Don't step on the grass"). In our theory they differ by their requirements of information inclusion (the former requires reference to the active goals and plans of the mind that generates the meaning, while the latter may not). We see these two areas of meaning as sharing a large number of features, which we will detail in the next sections.

As mentioned, our theory is about foundational meaning, so any related theory that does not address the question of how meaning is *generated* does not have direct bearing on the particulars of our work described here. This includes the well-known theories of Wittgenstein [17], Grice [6], and many others, whose focus on the meaning of language and symbols renders their theories (mostly) devoid of attempts to outline the particular mechanisms necessary for systems that generate their own meaning. Similar can be said about what have been called Referential theories of meaning [1,2,4,8]. Linguistic theories of meaning concerned with sound-meaning relationships are likewise orthogonally related to

[1] See http://www.openaera.org—*accessed Jun. 1st, 2024.*

efforts of building AI systems that generate meaning. An overview of a wide range of philosophical theories of meaning can be found in [5].

In addressing the concept of intelligence, consciousness, and thought, Dennett proposed the concept of the 'intentional stance' [3], which can be adopted by any third party wanting to explain the behavior of an observed complex system that is goal-directed and is expected to act (or aim to act) rationally. This view is very compatible with our theory; one could say that ours begins where it ends.

Our approach to the subject of meaning is in line with Peirce's Pragmatic Maxim [10,11], which asserts that the meaning of a concept resides in its conceivable practical effects or consequences and is revealed through its potential impact on experience and behaviour.

3 Definitions and Concepts

A **world** W is a formal description of a set of constraints and processes having its own universal clock that establish universal ground truth for an intelligent agent. W consists of a set of variables $V = \{v_1, v_2, ..., v_{||V||}\}$, a set of dynamic functions F, an initial state S_0, and a set of relations $\Re$ between the variables, formally $W = \langle V, F, S_0, \Re \rangle$. The variables represent everything that may change or hold a particular value in the world. The dynamic functions define the *laws of nature* in W and update its *full state*: $S_{t+\delta} = F(S_t)$ for each $t + \delta$ timestep. The dynamic functions consist of a set of transition functions $F = \{f_1, f_2, ..., f_n\}$ where $f_i : S^- \rightarrow S'^-$ and S^-, S'^- are partial states. *Invariant relations* in W are conditions or properties over the variables of W that remain unchanged as W transitions from one state to another. W is assumed to be nondeterministic, otherwise the concept of 'choice' would be empty (see below). W's clock limits measurements and intervention to a moment in time (the "now" – a demarcated interval), anchoring both to a particular moment (or interval) in time, in relation to other events. Measurements and manipulation of W can only be done in the *now*. No changes can be made to the past; measurements of the past can be made under the assumption of no intervening changes having happened; measurements and changes of the future can be planned through prediction (see below).

A world **state** S is a set of variables in the world, $S \subset W$, that have assigned values (some of which can be measured by an agent through its sensors, and affected through actuators). A **situation** σ refers to the immediate surroundings of an agent, where a subset of a world (a set of states) can be relatively easily measured and affected by the agent at any given time, $\sigma \in W_t$.

An **agent** is an embodied system consisting of sensors and actuators, and a controller (the mind), implemented in some computational substrate (all together this constitutes the agent's *body*). The controller's substrate contains resources (compute power and knowledge) that can be committed to cognitive tasks by an attention process. The body defines the agent's interface to the world, which allows the measuring (perception) of variables in W, through the flow of energy from the body's sensors to the controller, turning it into data, and the flow of energy towards its end-effectors for the execution of atomic actions,

by means of commands initiated by the controller for the body's actuators. An agent can be assigned a (set of) top-level drive(s), which define its reason for existence (in the sense that a vacuum cleaner's purpose is to keep the floors clean); all goals and subgoals are derived from this top level (cf. [13]).

A **controller** consists of a set of processes P that can receive an input, $i \in I$, produced by either measuring the world W or by producing reflectable knowledge (inspectable and dissectable thoughts), current state S, at least one goal $g_x \in G$ (implicit or explicit) and output $o \in O$ in the form of atomic actions (selected from a set of atomic possible outputs O), that (in the limit) achieve goal(s) G. We define **autonomy** not as the size of the space of actions that are available to the agent at any point in time, but rather, the ability of an agent to act based only on its own knowledge, without having to 'call home' (for the intervention of an external controller, e.g. a designer or teacher).

An agent $\mathcal{A}$ is **situated** if $\mathcal{A}$ is embodied and positioned in a particular circumstance or situation σ that is subject to limited energy, space and time (LEST). Our theory of meaning pertains only to the physical world (making no claims about abstract or hypothetical ones); a situated agent in our theory is thus always subject to LEST. A situated agent's ***action potential*** is its potential to assert changes on the environment, in the pursuit of its active goals, as limited by the current situation (through LEST).

A **goal** state (positive goal) is a desirable (possibly partially defined) state that the agent could or should reach. Conversely, a failure state (negative goal) is an undesirable state that the agent should avoid. The goals an agent is ultimately expected to pursue are called top-level goals (G_{top}) and derive from $\mathcal{A}$'s drives (see above). A goal can be decomposed into a number of subgoals, which describe and constrain how it can be achieved, resulting in a goal hierarchy with top-level goals at the top and atomic actions at the bottom. At any given time, only a subset of the goal hierarchy is commmitted to being pursued – the **active goals**. A **plan** is a sequence of (atomic) actions extending over a period of time, whose successful execution is expected to update the state of the world according to the agent's goals and subgoals, so that a goal (or subgoal) state is achieved and/or a failure state is avoided. To realize plans, the agent's runtime operation involves generating, evaluating, replacing and committing to goals, while ensuring sensible usage of available resources and knowledge, responsiveness to unexpected events, in a mixed planning/opportunistic manner.

A learning agent's knowledge $\mathcal{K}$ consists of a growing and changing set of goals and **models**, endogenously formed from its experience. Good models allow an agent to systematically predict, affect, explain and re-create phenomena [14]. To do so effectively and efficiently, models must capture **causal relations** between (hypothesized) causes and their effects [15]. The better (effective, efficient, useful) the models are for these purposes, the more **reliable** they are.[2] Causal models describe the evolution of a substate of the world $S \subset W$ as a conditional state

[2] Models of the physical world generated by an embodied agent may always turn out to be incorrect; guarantees of 'truthfulness' cannot be given. Models thus cannot be said to be 'correct' or 'true,' only *useful* and *reliable*.

transition function (the fewer conditions, the more general it is) that applies a (conditional) rule R to S producing the new state S': $S \xrightarrow{R} S'$. Causal models can be combined in various ways, via ampliative reasoning [13], to describe the transformation of one world state S to another S', in e.g. the attainment of goals and whether they are realistically possible or not (one role of intelligence is figuring out which ones are). The application of N causal models, where the input state of each model m is the output state of the previous model $m-1$, forms a *causal chain* of N sub-states potentially reachable by the system. Causal chains are built through a process that applies (non-axiomatic, defeasible) reasoning to models, in the service of goals.

A situated agent's knowledge about a particular situation is **grounded** if an unbroken contextualized chain of causal models can be formed that connects the low-level perceptual data, generated in the 'now' from the agent's situation, to its top-level goals. A chain is "broken" if the reliability of any of the causal models forming the chain is below a minimum threshold; this reliability is explicit metaknowledge that describes the chain itself that can be reasoned over (i.e. supports reflection) like models and goals. To consistently pursue and achieve goals, predict and model its situation, $\mathcal{A}$ must have some minimal set of causal and relational models of the tuple $\langle \mathcal{A}, \sigma \rangle$. The agent's model of self is built in the same way as other knowledge, consisting of models of its own mind and body, initially derived from its seed (the knowledge the agent was born with) and progressively refined, expanded and generalized through experience.

Prediction is an ongoing cognitive process Pr that produces hypothetical future states S that can be used as the basis for generating new goals G and plans Pl. Pr implements (non-axiomatic, defeasible) abduction and (non-axiomatic, defeasible) deduction processes, both of which rely on models of cause-effect relations. Due to physical limitations in cognitive resources, an agent must select which predictions to make for any period of time.

Uncertainty exists in all knowledge, stemming from imprecision in *measurement* (because taking reliable measurements takes time, and time is always limited), *predictions* (because they are produced from defeasible knowledge), and *control* (due to unknown factors – 'noise'). A **choice** is made by an agent through a commitment of resources to one option from a set of mutually exclusive alternatives. This includes both thinking and interacting with the world.

4 Meaning Generation

Now we can outline how foundational meaning is generated in an autonomous situated agent. The following applies equally to agents found in nature and those manufactured in a lab.

Definition. The *foundational meaning* of a datum $\mathcal{I}$ to agent $\mathcal{A}$ in situation σ is constituted by a causal coupling of $\mathcal{A}$'s (model of its) *situated future* to (its models of) its own action potential – what it considers itself to be capable and not capable of doing, in light of $\mathcal{I}$, in the form of (active and non-active) goals and plans, represented in an explicit hierarchy of relevant knowledge $\mathcal{K}$ – such that, based on $\mathcal{A}$'s active goals $\mathcal{G}$ and resulting new goals G and plans Pl, any relations of $\mathcal{I}$'s to $\mathcal{A}$'s knowledge can be (causally) traced to $\mathcal{A}$'s top-level goal(s), G_{top}.

By 'datum' we mean an information structure containing parts that are recognizable (at least to some minimal extent) by an agent. The reason it must have recognizable (classifiable) parts is that without any such features, the datum cannot be processed by the agent's cognition, and would thus be meaningless ($\mathcal{I}$ would be 'incogitable' – invisible to the agent's cognitive mechanisms).

Causal knowledge is a necessary requirement for systematic, efficient and effective control of a phenomenon. An agent $\mathcal{A}$'s *situated future* consists of its *predictions of what may and may not happen in its spatio-temporal proximity*, given its knowledge of cause-effect relations relevant to σ. The causal coupling centers around causal relations, and related relevant knowledge, that form an unbroken (non-axiomatic, defeasible) chain that connects top-level goals and low-level perceptions, contextualizing $\mathcal{I}$ in $\mathcal{A}$'s knowledge network. The more reliable the weakest causal link is in this chain, the *stronger is the grounding* of the meaning of $\mathcal{I}$ to $\mathcal{A}$. $\mathcal{A}$'s *action potential* is the set of actions (including measurements) that can be performed by $\mathcal{A}$ with respect to $\mathcal{I}$ in a given situation σ (σ defines which variables are observable and manipulable by $\mathcal{A}$ and σ thus constrains $\mathcal{A}$'s action potential). We refer to this overall information content as an "$\mathcal{M}$-structure."

Creating an $\mathcal{M}$-structure involves reasoning (non-axiomatic, defeasible deduction, abduction, induction and analogy – in a variety of combinations). Such reasoning also determines whether the chain's strength, whether it is unbroken, and the whether, and how, certain causal models of the world subsume others. Reasoning is also necessary to figure out implications of $\mathcal{I}$ in the present situation σ, which unavoidably consists of models of relations between diverse relevant components – sensors, actuators, objects in the environment, forces, obstacles, etc. – forming a hierarchy of goals, models, plans, parts and wholes.[3] The requirement for manipulable models of whole-part relations and constraints means that the knowledge must be (partly or fully) symbolic;[4] the existence of

[3] For instance, a shoe is made up by parts like laces, soles, etc.; causal relations define what can and cannot be done with the shoe and its parts in particular situations, given certain constraints (e.g. whether we are wearing them or just looking at them).

[4] For practical reasons, including memory storage and compute power, there will always exists a level of detail below which relevant information will not be strictly symbolic, and another lower one below which it will be completely sub-symbolic. This is most obvious for low-level perceptual data, e.g. vision, which may involve bandwidths of 2 megabits per second or more.

novelty and uncertainty in a non-axiomatic world means that the process and runtime of the reasoning cannot be known beforehand, so it must be ampliative and learnable.[5] Learning to do ampliative reasoning, in turn, cannot be done without reflection over both the domain knowledge and the reasoning mechanisms themselves, which calls for a transparent compositional explicit representational scheme.

To ensure a certain level of alertness and action capacity, a situated agent must generate meaning continuously. Meaning generation is a situated thought process that is linked to the 'now' via active goals $\mathcal{G}_t$, where t is a (short) interval. In other words, grounding – grounded cognition – happens via linking a situation σ_t to active goals $\mathcal{G}_t$ using (reliable) causal models.

The **foundational meaning** $\mathcal{M}$ of $\mathcal{I}$ for agent $\mathcal{A}$ in situation σ is thus captured by the $\mathcal{M}$-structure:

$$\mathcal{M}^{\mathcal{A}}_{now}(\mathcal{I}) = Pr_t(\mathcal{I}, \sigma_t, \mathcal{K}) \rightarrow G_{t'} \rightarrow Pl_{t''} \quad (1)$$

where: *now*: *minimum time interval for meaning generation; now* $= t'' - t$
$\mathcal{I}$: *information structure* $\mathcal{A}$: *agent (embodied controller)*
Pr: *predictions* σ: *situation as modeled by the agent*
$\mathcal{K}$: *the agent's prior knowledge deemed relevant to* $\mathcal{I}$ *and* σ
$G_{t'}$: *new or updated goals (active and/or passive) created from* Pr_t
Pl: *new plans created from* $G_{t'}$

'Now' spans the time from the measurement when $\mathcal{I}_t$ is received, to the time that a plan $Pl_{t''}$ has been produced from the goals $G_{t'}$, that in turn derive from the predictions Pr_t.

Question: *Why are predictions part of the* $\mathcal{M}$*-structure?*

An intelligent agent will always be in pursuit of a set of goals, in the very least due to the constantly changing world it's situated in. The resources needed to achieve goals, including time, energy, space, and knowledge, the situation's 'now' puts constraints on their achievement. An implicit goal of an intelligent agent is keeping its action potential high (maximizing its potential choices). When a controller is presented with a new piece of information, $\mathcal{I}$, the relation of this information to its active goals must be assessed, to see how the action potential for them may be impacted. This impact is always in the future; predictions and plans are a way to detail the shape of such impact. Pursuit of impossible goals is a waste of a controller's resources, so it must be capable of making reasonable predictions about which goals are possible and more sensible than others, and why. These predictions must say something about the future state of the situation σ and the agent itself – including its knowledge, embodiment, and presently active goals $\mathcal{G}$. If an important active goal, e.g. staying alive, is predicted to be heavily impacted or categorically prevented, this has profound implications for the agent's existence (including the fact that all of its other

[5] In our approach, ampliative reasoning' includes (non-axiomatic) deduction, abduction, induction and analogy. These must be dynamically chosen at runtime based on situation and active goals.

active goals will also be prevented). Meaning in a thinking system is thus carried by predictions.

▷ *Predictions participate in the $\mathcal{M}$-structure by ensuring the reliability of committed goals and plans.*

Q: *How does an agent's perception of a situation affect meaning generation?*

The current situation, including an agent's embodiment, constrains what the agent can sense and affect at any particular point in time – and this is critical for determining what knowledge to use for producing goals and plans. Creating useful (new) models is also dependent on perception and classification of the present context, and models are the smallest relevant unit for generating meaning, supporting prediction and action. Predictions that are not grounded in the 'now' and don't take into account recent changes in the world will be based on outdated information, resulting in incorrect predictions, which in turn will produce invalid goals and plans.

▷ *The agent's model of the current situation is part of the $\mathcal{M}$-structure.*

Q: *How does an agent's prior knowledge participate in determining meaning?*

To make sense of new information $\mathcal{I}$, an agent has to relate it to its knowledge, including its active goals. This includes dissecting $\mathcal{I}$ into its constituent components and classifying the present situation σ. The only way to do either is by using existing knowledge. An agent's top-level goals incarnate its mission; any active subgoals will by definition be related to this mission. Information that may help or hinder it in pursuing its active goals will thus be relevant to its existence. For example, if someone points out a pedestrian to an agent driving a car, extracting their walking direction may predict them to be on a direct collision course with the car. The classification in this example requires knowledge about the look and behavior of pedestrians, their mode of locomotion and speed, and the same for the controlled vehicle. Avoiding harm to others may be one of the driver's negative goals. The meaning generated from this dissection – the potential for causing future harm – thus xaffects an important top-level goal of the agent. There is no condition under which such classification could be done without some knowledge about the phenomena in question, at various levels of detail, in light of the goal.

▷ *The agent's prior knowledge is part of the $\mathcal{M}$-structure.*

Q: *Why are new/updated goals part of meaning generation?*

Full autonomy requires a system to generate goals and subgoals on its own. At any moment in time, only a subset of the system's goals are active; active goals can be seen as a set of (multi-dimensional) attractors (consisting of subgoals) that steer the agent's behavior, with the intent of transforming a situation to align with its goals. The autonomous goal-generation process is driven by predictions and reasoning over models of the current knowledge of the situation. Any perturbation of a path towards a goal may require changing some subset of a plan or re-planning from scratch. Predicted perturbations may also require a goal to check at a later time whether the prediction came true. In either case, the end-product of such efforts would be described by a new or updated goal.

The meaning is carried not only by the predictions of future states of the present situation but also the relation of these predictions to the presently active goals and, ultimately, their relation to the agent's top-level goals. Hence, updated and new goals are a necessary step in the meaning generation.

▷ *Updated/new goals are part of the $\mathcal{M}$-structure to ensure an agent's mission.*

Q: *Why are plans part of meaning generation?*

Any situation's meaning – to the system itself – is captured by the system's predictions of how the situation develops into the future, and how this development affects the currently active goals. Given the predicted effect of the present situation σ_t on the active goals $\mathcal{G}_t$, and adjusted goals $G_{t'}$ that take those predictions into account, new plans Pl_t provide actionable descriptions of how these could be achieved. The generation of plans, and their form, is a necessary part of the meaning generation process because it outlines what is possible and what is impossible under the constraints presented by σ and $\mathcal{I}$. Plans thus provide a mechanism for distinguishing between useful and useless (meaning*ful* and meaning*less*) information with respect to active goals and the situation, in light of available knowledge and resources. This, then, is the "end of the line" of the meaning generation process, feeding back to new predictions, which then produces new updates to the present meaning structure.

▷ *Plans are part of the $\mathcal{M}$-structure by clarifying the importance of information.*

Q: *Why is meaning always bound in the 'now'?*

Because the physical world has an immutable 'ticking clock.' Thinking is control, and control is about making choices in light of options, under the constraints of the world clock. The past cannot be changed, and only choices made in the 'now' can affect the future – all agents in the physical world are prisoners of the 'now.' Actions are thus only to be enacted in the 'now.' Simply by 'ticking,' the world clock changes the action potential of all agents, independently of what they do. This means that the foundational meaning of an agent's knowledge will change eventually. Meaning generation is an implemented (physical, "cyber-physical") computational process. Computation cannot happen in the past or in the future: for an autonomous controller the 'now' thus serves as the anchor point for interpreting the current situation; in the 'now,' fixated at a particular point in time, because for meaning to exist, the outlined computations must be performed. In other words, the generation of meaning is a dynamic process that has to be executed by a concrete (physical) computing agent. Regardless of previous events, meaning depends only on the information and knowledge currently available to an agent. Semantic meaning thus depends on foundational meaning.

▷ *Semantic and foundational meaning are always anchored in the 'now.'*

Q: *Is there a difference between meaning and autonomy?*

What sets intelligence apart from all other phenomena, and thus defines it, is its ability to handle novelty [16]. Meaning is the causal linking of novelty, knowledge, a situation in the now, predictions, goals and plans. Meaning generation involves the effective and efficient outcome of this process – we say that

an agent that can do this reliably and repeatedly is able to extract the meaning of situations. This meaning extraction capability allows an agent, in turn, to act autonomously. Meaning generation is not needed if autonomy is not desired or needed; without autonomy there is no need to generate meaning. Autonomy and meaning are in fact two sides of the same coin: *Autonomy without meaning generation is meaningless; meaning generation without autonomy is pointless.*
▷ *Autonomy ≡ Meaning.*

5 Conclusions and Future Work

Meaning and meaning generation are key aspects of highly autonomous, generally intelligent systems. This work contributes to bridging a gap, for too long left open, between theories of meaning and the design of systems that can generate and handle meaning. Our theory identifies precisely the qualities necessary for a system to generate meaning for itself; the hope is to pave the way for a new class of intelligent systems showing unprecedented autonomy and generality. A more exhaustive treatment of the implications of this theory, restricted here for reasons of space, is deferred to future work.

Acknowledgment. The authors would like to thank the GMI team at Reykjavik University, Bridget Burger and the anonymous reviewers for insights and discussions.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this paper.

References

1. Alston, W.P.: Meaning. In: Edwards, P. (ed.) The Encyclopedia of Philosophy, pp. 5–233. Macmillan (1967)
2. Braun, D.: Russellianism and explanation. Philos. Perspect. **15**, 253–289 (2001)
3. Dennett, D.: Brainstorms. M.I.T. Press, Cambridge, MA (1978)
4. Frege, G.: Sense and reference. Philos. Rev. **57**(3), 209–230 (1948)
5. Gelepithis, P.: Survey of theories of meaning. Cogn. Syst., 141–162 (1988)
6. Grice, H.: Studies in the Way of Words. Harvard University Press (1989)
7. Lewis, D.K.: General semantics. Synthese **22**(1–2), 18–67 (1970)
8. Lyons, J.: Language, Meaning, and Context. Fontana, [London] (1981)
9. Nivel, E., Thórisson, K.R., Dindo, H., Pezzulo, G., et al.: Autocatalytic endogenous reflective architecture. Technical RUTR-SCS13002, Reykjavik University School of Computer Science, Reykjavik, Iceland (2013)
10. Peirce, C.S.: Pragmatism as a Principle and Method of Right Thinking: The 1903 Harvard Lectures on Pragmatism. State University of New York Press (1997). Turrisi, P.A.: (ed.)
11. Peirce, C.S., de Waal, C.: Illustrations of the Logic of Science. Open Court, Chicago, Illinois (2014)

12. Prior, A.N.: Correspondence theory of truth. In: Edwards, P. (ed.) The Encyclopedia of Philosophy, vol. 2, pp. 223–224. Macmillan (1967)
13. Thórisson, K.R.: Seed-programmed autonomous general learning. Proc. Mach. Learn. Res. **131**, 32–70 (2020)
14. Thórisson, K.R., Kremelberg, D., Steunebrink, B.R., Nivel, E.: About understanding. In: Steunebrink, B., Wang, P., Goertzel, B. (eds.) AGI -2016. LNCS (LNAI), vol. 9782, pp. 106–117. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41649-6_11
15. Thórisson, K.R., Talbot, A.: Cumulative learning with causal-relational models. In: Iklé, M., Franz, A., Rzepka, R., Goertzel, B. (eds.) AGI 2018. LNCS (LNAI), vol. 10999, pp. 227–237. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-97676-1_22
16. Wang, P.: On defining artificial intelligence. J. Artif. General Intell. **10**, 1–37 (2019)
17. Wittgenstein, L.: Philosophical Investigations. Basil Blackwell, Oxford (1953)