Cognitive Systems: Argument and Cognition

Antonis Kakas , Loizos Michael

***Abstract*—Developing systems that are aware of, and accommodate for, the cognitive capabilities and limitations of human users is emerging as a key characteristic of a new paradigm of cognitive computing in Artificial Intelligence. According to this paradigm, the behavior of such *cognitive systems* is modeled on the behavior of human personal assistants, able to understand the motivations and personal likings / affinities of their interlocutors, while also being able to explain, and ultimately persuade the latter about, their computed solution (e.g., a proposed action) to a problem.**

This paper examines the link between argument and cognition from the psychological and the computational perspectives, and investigates how the synthesis of work on reasoning and narrative text comprehension from Cognitive Psychology and of work on computational argumentation from AI can offer a scientifically sound and pragmatic basis for building human-aware cognitive systems for everyday tasks. The paper aims, thus, to reveal how argumentation can form the *science of common sense thought* on which new forms of cognitive systems can be engineered.

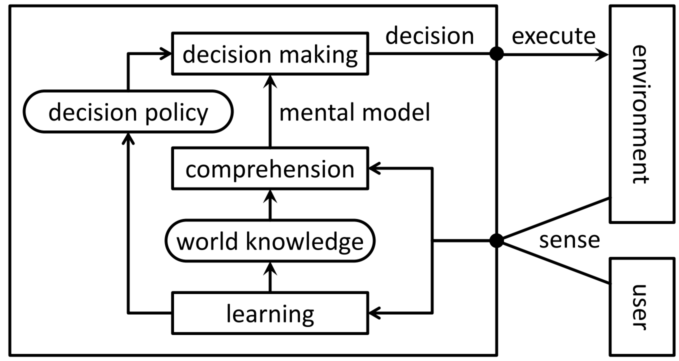

Fig. 1. High-level view of the architecture of a cognitive assistant, with focus on the interaction of the processes of decision making, comprehension, and learning. The components labeled as "decision policy" and "world knowledge" correspond, respectively, to the sets of option arguments and belief arguments.

I. The Emerging Need for Cognitive Systems

THE ever increasing demand for smart devices with ordinary human-level intelligence, capable of common sense reasoning and attuned to everyday problem solving, is forcing Artificial Intelligence to stand up and deliver. Unlike anything seen to date, this new vision of user-device interaction aims to allow ordinary users, without technical background, to instruct or program their devices in a natural and personalized manner, and to allow the devices to assist (and enhance the abilities of) their users in dealing with everyday tasks. This *symbiotic* relation splits the burden of communication among the user and the device, giving rise to a "programming paradigm for the masses" [1] that avoids the extremes of using natural languages that are too complex for ordinary devices, or programming languages that are too complex for ordinary users.

Early examples of systems exhibiting such symbiotic interactions already exist, ranging from personal assistant software provided by major smart-device manufacturers, to the expected application of systems that extract information from massive amounts of unstructured data for the purposes of expert-level analysis of problems in specialized domains (e.g., health, law).

Unlike existing automated systems, these *cognitive systems* [2] often exhibit an operational behavior resembling that of a human personal assistant. In particular, a cognitive system's domain of application is limited to certain common everyday tasks, and its operation revolves around its interaction with its human user in a manner that is compatible with the cognitive reasoning capabilities of the latter. To understand (and correct when needed) the reasoning of the system, the user expects the system to use *common sense* to fill in any important relevant information that the user leaves unspecified, and to be able to keep learning about the domain of application and the user's personal preferences and beliefs through their interaction.

Efforts to meet this emerging need and to guide the future of cognitive systems is bound to benefit from a foundational basis that facilitates a human-device interaction that places cognitive compatibility with humans at the center stage. This paper puts forward computational argumentation as a candidate for this reconciliation between human and machine reasoning, in a manner that is more appropriate than the classical logic basis that underpins the development of automated systems to date.

II. Argumentative Basis of Human Cognition

Given the emphasis of cognitive systems on cognitive compatibility, an argumentative foundation for their development will be a viable option only if human cognition is itself geared towards an argumentative perspective. We overview work from Psychology that provides evidence in support of this condition.

A significant amount of research in the area of Psychology of Reasoning over the last century suggests that, in comparison with strict classical logic, human reasoning is failing at simple logical tasks, committing mistakes in probabilistic reasoning, and succumbing to irrational biases in decision making [3], [4]. Different interpretations and theories on the nature of human reasoning have been proposed to explain these findings. Certain proposals attempt to stay very close to the mathematical and strict form of logical reasoning, such as "The Psychology of Proof" theory [5], which proposes a psychological version of a proof system for human reasoning in the style of Natural Deduction. Despite its many criticisms (see, e.g., [6] for a thorough and critical review of this theory), the theory shows a necessary departure from the proof systems of classical logic.

This paper grew out of tutorials given at IJCAI 2016 and at ECAI 2016. Information on the tutorials is available at: http://cognition.ouc.ac.cy/argument/.

Antonis Kakas is a professor in the Department of Computer Science, University of Cyprus, Nicosia, Cyprus; e-mail antonis@ucy.ac.cy

Loizos Michael is an assistant professor in the School of Pure and Applied Sciences and director of the Computational Cognition Lab, Open University of Cyprus, Nicosia, Cyprus; e-mail loizos@ouc.ac.cy

More importantly, the theory implicitly indicates that human reasoning is linked to argumentation, since proof systems like Natural Deduction are known to have a natural argumentative interpretation [7]. Other proposals (see, e.g., [8]) completely abandon any logical form for human reasoning, treating it as the application of specialized procedures, invoked naturally depending on the situation in which people find themselves.

Earlier work demonstrated empirically that humans perform with significant variation in successfully drawing conclusions under different classical logic syllogisms [9]. The study of the Psychology of Syllogisms [10]–[12] proposes that humans use mental models to guide them into drawing inferences, which foregoes the "absolute and universal" validity of the inferences supported by reasoning based on truth in all possible models of the premises. Instead, a mental model captures reasoning based on the *intended interpretation* of the premises, and corresponds to a suitable situation model, much like what humans construct when processing or comprehending a narrative [13], [14].

In a modern manifestation of this perspective in the context of Computational Logic in AI [15], it is argued that structures like mental models are a useful way to capture various features of human reasoning, not least of which its defeasible nature. Building mental models can be seen as building arguments to support an intended interpretation of the evidence currently available, by combining them with general rules of common sense knowledge that people have acquired. The mental model approach to deduction can, then, be reconciled with the view of reasoning through inference rules, while the defeasible nature of reasoning follows from the defeasible nature of arguments.

In addition to the plethora of psychological findings that are consistent with, and indicative of, an argumentative interpretation of human reasoning, some more recent work from the Psychology of Reasoning provides further explicit evidence in support of this position [16]. Supported by the results of a variety of empirical psychological experiments, the authors of that work propose that human reasoning is a process whereby humans *provide reasons* to accept (or decline) a conclusion that was "raised" by some incoming inference of the human brain. The main function of human reasoning, then, is to lay out these inferences in detail, and to form possible arguments that will produce the final conclusion, in a way characterized by the awareness not just of the conclusion, but of an argument that justifies accepting that conclusion. Through the process of human reasoning, therefore, people become able to exchange arguments for assessing new claims, and the process of human reasoning becomes, effectively, a process of argumentation.

Experiments carried out to test how humans form, evaluate, and use arguments, suggest that humans produce "solid" arguments when motivated to do so; i.e., in an environment where their position is challenged. If unchallenged, the arguments initially produced can be rather naive, until counter-arguments or opposing positions are put forward, at which point humans produce better and well-justified arguments for their position by finding counter-arguments (i.e., defenses) to the challenges. For example, in experiments where mock jurors were asked to reach a verdict and then were presented with an alternative one, it was observed that almost all of them were able to very quickly find counter-arguments against the alternative verdict, while strengthening the arguments for their original verdict.

The experimental results indicate that automating human reasoning through argumentation can follow a model of computation that has an *"on-demand"* incremental nature. Such a model of computation is well-suited in a resource-bounded problem environment, and more generally for the development of cognitive systems under the personal assistant paradigm.

Overall, work from Psychology has exposed some *salient features* of human reasoning directly related to argumentation: *(i)* handling of contradictory information, by acknowledging the defeasible nature of knowledge; *(ii)* drawing of tentative conclusions, which are revised in the presence of more information; *(iii)* awareness not only of a conclusion, but also of its justification; *(iv)* "on demand" / dialectical reasoning that defends challenges as they arise; *(v)* use of a single intended mental model, while accommodating common and individual biases across humans. Collectively, these features suggest that argument is native to human reasoning, and, consequently, that argumentation can offer a unified perspective of empirical psychological evidence on the nature of human reasoning.

III. Computational Argumentation in AI

Efforts to formalize human reasoning in terms of an argumentation theory can be traced back to the work of Aristotle and his notion of "dialectic argument". Until rather recently, argumentation was primarily studied from a philosophical and / or a psychological perspective. These works [17]–[19] have helped generate a new interest on the study of argumentation within the field of AI, with motivation coming from both *(i)* the desire to have intelligent systems with human-like defeasible (or non-monotonic) reasoning, and belief revision capabilities in the face of new information [20], [21], as well as *(ii)* the study of the dialectic nature of reasoning in various areas of human thought, such as rhetoric and legal reasoning [22]–[26].

The early 1990s saw the introduction of *abstract argumentation* [27], where arguments are considered as formal entities separate from the particular context in which they arise, and are viewed only in terms of their syntactic and semantic relationships. This view emerged from work [28], [29] showing that argumentation could capture most of the existing non-monotonic logical frameworks, and, hence, provide a uniform way to view the aspect of defeasibility in human reasoning.

An abstract argumentation framework is defined as a tuple $\langle \mathcal{A}, \mathcal{R} \rangle$, where $\mathcal{A}$ is a finite set of arguments and $\mathcal{R}$ is a binary (partial) relation on $\mathcal{A}$, called the *attack relation* on $\mathcal{A}$. This attack relation is lifted to subsets of arguments, so that a subset A of $\mathcal{A}$ attacks another subset B of $\mathcal{A}$ if and only if there exists $a \in A$ and $b \in B$ such that a attacks b; i.e., $(a, b) \in \mathcal{R}$. One is then concerned with the problem of building "good quality" or *acceptable* argument subsets $\Delta \subseteq \mathcal{A}$ that "defend against" or attack back all possible argument subsets that attack Δ, and which constitute, therefore, counter-arguments to Δ.

A general way to formulate a notion of acceptability is the dialectical definition that an argument subset Δ is acceptable if and only if any argument subset A that attacks Δ is attacked back by some argument subset D (i.e., D defends Δ against A) that is, itself, "acceptable with respect to Δ". There are several different ways to offer a precise formulation of what is meant by the condition that D is acceptable with respect to Δ, such as: *(i)* that D is simply a subset of Δ, which gives rise to the admissibility semantics of argumentation, or *(ii)* that D is (eventually) an argument subset that is not attacked by any other argument subset (and is, hence, globally undisputed), which gives rise to the grounded semantics of argumentation.

This simple, yet powerful, formulation of argumentation has been used as the basis for the study and development of solutions for different types of problems in AI [30], [31]. In particular, it forms the foundation for a variety of problems in multi-agent systems (see, e.g., the workshop series "ArgMAS: Argumentation in Multi-Agent Systems") where agents need to exhibit human-like autonomy and adaptability. Recently, the area of *argument mining* (see, e.g., [32] for an overview) aims to provide an automatic way of analysing, in terms of formal argumentation frameworks, human debates in social media, even by identifying relations that are not explicit in text [33].

In many of the application domains above, a realization of abstract argumentation is used where the attacking relation is materialized through a priority or preference relation between conflicting arguments. Such *preference-based argumentation* frameworks consider more preferred arguments to be stronger than, and thus to attack, less preferred arguments, but not vice-versa. Preferences can be derived naturally from the particular domain of application, capturing general or contextual aspects of the domain, or biases and beliefs of individual agents.

More recently it has been shown that even classical logical reasoning, as found in formal mathematics, can be captured in terms of abstract argumentation [7]. In such an *argumentation-based logic*, logical entailment of some conclusion is obtained through the existence of an acceptable argument supporting the conclusion and the absence of acceptable arguments that support any contrary conclusion. This suggests that argumentation need not be approached as a substitute for classical logic, but as an extension thereof that is appropriate for reasoning both with consistent premises but also with inconsistent ones.

The aforementioned studies of argumentation in AI show that computational argumentation has the capacity to address the salient features of human reasoning that have been pointed out by empirical psychological studies. Argumentation offers a natural form of reasoning with contradictory information, by supporting arguments for conflicting conclusions, and handling the retraction of tentative conclusions whenever new stronger arguments emerge. Furthermore, argumentation gives a form of reasoning that is based on an intended mental model that comprises the conclusions that are supported by the strongest available arguments, and provides explicit justifications in support of that intended model. Lastly, argumentation explicitly adopts "on demand" reasoning through a dialectical definition of acceptability, while its preference-based realization readily accommodates for human biases and individual beliefs.

IV. Argument and Human Decision Making

Having offered evidence for the capacity of computational argumentation to capture the salient features of human reasoning, we turn our attention to how argumentation can be utilized in the development of cognitively-compatible systems.

An important subclass of cognitive systems will be that of *cognitive assistants* that help their human users take decisions in everyday tasks: which restaurant to visit for some occasion, when to schedule a meeting, or how to handle an information overload on a social network. Despite appearing relatively simple when compared with the complex optimization problems that conventional computing systems solve, these everyday decision-making problems come with their own challenges.

Any systematic and principled attempt at developing cognitive assistants needs to account for several characteristics of human decision-making that have been exposed by work in Cognitive Psychology: departure from the formal decision theory and influence by biases (e.g., earliest information, similar past decisions, group conformity), consideration of individual preferences and predispositions, minimal initial evaluation of the options and additional evaluation as the need arises.

Beyond ensuring cognitive compatibility, one must also account for pragmatic considerations. Decision-making is rarely an isolated process, and the arrival of extra or revised information in an *open and dynamic environment* may affect decision-making by: offering new options (e.g., a new restaurant just opened up); rendering existing options (physically) inapplicable (e.g., the boss cannot meet after 11:00am); or revealing updated values for options (e.g., an online community that the user had enjoyed following started using offensive language).

The *challenge of building cognitive assistants* resides, thus, in being able to coherently operate at three levels: *(L1)* represent information akin to the user's general motivations and desires, and the system's beliefs of the state of the world at the time when decisions will be effected; *(L2)* offer explanations / justifications of the proposed decision that are meaningful to the user; *(L3)* participate in a dialectic debate process to either persuade the user of the proposed decision, or to revise it.

The natural solution that argumentation offers for level *(L2)* and level *(L3)* has been utilized in several works in AI dealing with decision-making in contexts ranging from legal decisions to informal human-like decisions by autonomous agents [34]–[39]. These works generally fall within the particular realization of preference-based argumentation, which also points to how argumentation can offer a solution for level *(L1)*, as well.

In an argumentation framework for a certain decision problem, each option is supported by one or more arguments. The structure of these *option arguments* can be represented simply by a tuple $\langle opt, val, bel\rangle$, where *opt* is the supported option, *val* is a set of user values (e.g., needs, motivations, desires) that the option and / or the argument serve, and *bel* is a set of beliefs that ground the argument on some information about the external world, in a manner that the cognitive assistant believes render option *opt* a possible alternative for consideration.

The values served by an argument can give a relative preference between arguments that reflects the *personal affinity* or interests that a cognitive assistant might be designed to follow. Thus, a preference between arguments $a_i = \langle opt_i, val_i, bel_i\rangle$ and $a_j = \langle opt_j, val_j, bel_j\rangle$ can be defined through the general schema that "a_i is preferred over a_j if $val_i \sqsupset val_j$", where $\sqsupset$ is a comparison ordering amongst the different values that can be based on the personality of the human user of the cognitive assistant. By concentrating on different values or on different comparison orderings, this simple general schema can give rise to different preferences over the same arguments, reflecting the natural variability across different contexts or different users.

In practice, human users may know heuristically from their earlier experiences the result of the evaluation of different arguments in certain situations, and hence that certain arguments are preferred over others. For example, a vegetarian may know that when having dinner outside her house, the vegetarian restaurant down town serves better her need to have a larger variety of choices, but the local market across the street offers a cheaper and faster choice. Instead of having to recompute her preferences based on her values for the two options, she might choose to simply state that when she is in a situation $S_{i,j}$ where she is currently at home and she has not eaten out during the past week, then she prefers the argument supporting the local market over the argument supporting the vegetarian restaurant, using the general scheme "a_i is preferred over a_j if $S_{i,j}$", where effectively $S_{i,j}$ is a situation in which $val_i \sqsupset val_j$.

The attack relation can be naturally defined from the preferences as follows: argument a_i attacks argument a_j if they support conflicting options, and a_i is not less preferred than a_j. Now, given a state S of the world, one can compute (under a chosen argumentation semantics) the acceptable arguments among those whose beliefs are compatible with S. Any option supported by an acceptable argument is a *possible* decision in S. Furthermore, a possible decision in S is a *clear* decision in S if there exist no other conflicting possible decisions in S.

To complete the picture, one must fix the choice of argumentation semantics. Given the nature of human decision-making, the natural choice for this case, and for the respective cognitive assistants, is that of the *grounded extension* semantics, which, as already discussed in the preceding section, can be derived as a special case of the dialectical definition of acceptability.

In summary, argumentation serves well as a basis for cognitive assistants that support human decision-making, offering natural solutions: *(L1)* at the representation level through the encoding of user-specific preferences and biases; *(L2)* at the decision-formation level through the incremental construction of acceptable arguments; *(L3)* at the persuasion level through the dialectic process of defending against alternative decisions.

V. Argument and Narrative Comprehension

In describing how abstract argumentation can be instantiated to support human decision-making, we have focused primarily on the role that values play in capturing the user's general motivations and desires, and have mostly side-stepped the role that beliefs play in capturing the applicability of arguments.

In the simplest case, these beliefs could be such that their compatibility against a given state of the world can be directly checked, outside the actual process of argumentation. More generally, though, the beliefs themselves are the outcome of a reasoning process, which could itself be argumentative. Thus, in addition to option arguments, the argumentation framework may also include *belief arguments*, supporting beliefs on which the option arguments rest. An option argument could, then, be potentially undercut by a belief argument that supports that the environment will not be conducive to the realization of the particular option, while a second belief argument that disputes this claim could be used to defend the option argument.

Exactly analogously to option arguments, belief arguments are evaluated against each other by means of a preference relation, which ultimately determines the attack relation between arguments. Unlike the typically user-specific preferences over option arguments, however, preferences over belief arguments naturally capture certain pragmatic considerations. These considerations derive primarily from the open and dynamic nature of the environment, which necessitates a cognitive assistant able to reason about missing information, the causal effects of actions, the passage of time, and the typical and exceptional states of the world. Belief arguments capture, then, knowledge about these aspects of the world, while preferences over belief arguments capture the commonsensical reasoning pattern that humans use to form a coherent understanding of the situation.

This type of reasoning is directly related to the process of narrative comprehension, with the coherent understanding of the situation corresponding to the intended interpretation of the narrative. During narrative comprehension, humans include in the intended interpretation information that is not explicitly present in the narrative but follows from it, explanations of why things happened as they did, links between seemingly unconnected events, and predictions of how things will evolve.

Starting with the seminal works of the Situation and Event Calculi, work in AI sought to codify the commonsense laws associated with reasoning about actions and change (RAC) in a narrative context, in terms of central problems to be solved: the frame problem of how information persists, by default, across time; the ramification problem of how actions give rise to indirect effects; the qualification problem of how action effects are blocked from materializing; the state default problem of how the world is not, by default, in some exceptional state.

Several works in AI [40]–[43] have demonstrated the natural fit of argumentation for RAC, by capturing the relevant aspects of human reasoning in terms of persistence, causal, and default property arguments, along with a natural preference relation between these different types of arguments. For example, a preference of causal arguments over conflicting persistence arguments cleanly addresses the frame problem by capturing the commonsense law of inertia that situations / information persist unless caused to change. Grounding the different types of arguments on information explicitly given in the narrative allows one to offer explanations for / against drawing certain conclusions at certain time-points or situations in the world.

Recent efforts [44] to combine an argumentation approach to RAC with empirical knowhow and theoretical models from Cognitive Psychology have led to the development of automated comprehension systems [45] that use belief arguments (under the grounded extension semantics, which we have proposed as appropriate for decision-making as well) to construct an intended mental model for a narrative, and appropriately update and maintain it in the presence of surprises and twists as the narrative unfolds. This treatment is not unlike what a cognitive assistant is expected to adopt when reasoning about its beliefs while the state of its environment unfolds over time.

Despite their predominant use to represent knowledge about the environment, belief arguments used by a cognitive assistant cannot be decoupled from its human user. The vocabulary and terms employed to express belief arguments should be familiar, their complexity should be manageable, and the justifications they give rise to should be meaningful, all with respect to the user. For example, a cognitive assistant's appeal to the belief argument that "pyrexia is not a medical emergency" might be inappropriate if its user is not familiar with the term "pyrexia" (fever), or if its user has been close to swamps (in which case fever might be indicative of malaria). These issues tie directly back to the requirement that cognitive assistants should operate in a manner that is cognitively-compatible with their users.

In summary, the argumentation basis for human decision-making, as proposed in the preceding section, can be naturally extended to address, *within a single unified framework*, the related and complementary problem of narrative comprehension: *(L1)* at the representation level through the encoding of world knowledge; *(L2)* at the decision-formation level through the construction of justifications that use concepts meaningful to the user; *(L3)* at the persuasion level through the grounding / contextualization of decisions on the fluctuating world state.

VI. Populating the Argumentation Arena

The acceptability semantics of computational argumentation can be effectively viewed as an *internal evaluation mechanism* for the quality of the conclusions of a cognitive assistant, with conclusions that are supported by stronger or more preferred (subsets of) arguments considered as being more pertinent than alternatives. Argumentation, however, does not provide for an analogous *external evaluation mechanism* for the quality of the cognitive assistant's arguments and preferences *in relation to* the environment. Equivalently, an argumentation framework is assumed to be populated with arguments and preferences of high external quality, and the acceptability semantics concentrates on the task of how to make a meaningful use of those.

A central means to populate an argumentation framework is through *cognitive programming* [1]. The user can volunteer, either during an initial familiarization period, or dialectically in response to a failure to be persuaded by the cognitive assistant, additional belief or option arguments and corresponding preferences, so that the cognitive assistant gradually becomes more "knowledgeable" about its environment, and better reflects the motivations and interests of its user. The requirement that a cognitive assistant's representation is cognitively-compatible with humans is key during this process, as the user naturally interacts with its cognitive assistant through the use of high-level concepts, and in a way that avoids detailed instructions.

More passively, the user may simply decline suggestions of the cognitive assistant without offering explanations / counter-arguments. Some form of online supervised learning can then be invoked by the cognitive assistant, with the user's feedback taken as a negative learning instance that the arguments supporting a decision are not acceptable, and that the preferences among arguments need to be revised to account for this. Under certain assumptions, the user's preferences between competing option arguments have been shown to be learnable [46], [47].

As an example of cognitive programming, the suggestion of a cognitive assistant to schedule a meeting of its user with his boss at 7:30am can be met by the user's response "Do not schedule work appointments too early in the morning.", which will thereafter be viewed as an extra argument, more preferred than the acceptable arguments that supported the suggestion. The importance of forming an intended model of the situation, and of the ability to employ common sense, is highlighted in this example, as the cognitive assistant needs to make sense of terms like "work appointment" and "too early in the morning".

Certain general belief arguments (e.g., that most people do not usually work during the nighttime) can be more reasonably acquired directly by the cognitive assistant, through manually-engineered or croudsourced knowledge-bases [48]–[51], and through the use of machine learning on text corpora [52]–[54]. A number of issues would, of course, have to be dealt with: the possible biases in the learning material, especially for text found on the Web [55]; the eventual use of the arguments by a reasoning process, without nullifying their learning-derived guarantees [56]–[59]; the availability of appropriate learning material to also acquire causal arguments [60]; the inadvertent effects of decisions supported by arguments that were learned outside the environment of the cognitive assistant [61], [62].

The autonomous or crowdsourced acquisition of arguments and preferences still benefits from user interaction. A cognitive assistant used for appointment scheduling, for example, typically has no immediate need for learning about forest fires. The user (or the manufacturer) can specify, then, relevant keywords to guide the cognitive assistant's search for applicable learning material. Alternatively, such keywords can be identified by the cognitive assistant by gathering those that occur frequently in its user's queries, so that autonomous learning can be invoked "on demand". Once arguments and preferences are learned, the user may further correct any misconceptions that have been acquired due to biases in the sources of the learning material.

It is important to note here that none of the processes of populating an argumentation framework restricts the application of cognitive assistants to common sense domains only. A medical professional, for instance, could cognitively program a cognitive assistant with arguments for making diagnoses of illnesses based on observed symptoms. The cognitive assistant could also autonomously learn medical arguments by restricting its search for learning material to medical ontologies and journals. Through these arguments, then, the cognitive assistant would be able to explain its medical recommendations in the same fashion that one medical professional would explain to another.

We conclude by observing that argumentation is not simply amenable to a process of learning, but rather a *natural fit* for it. Learned knowledge, especially when acquired autonomously by a cognitive assistant, cannot be strict, but can express only typical and defeasible relationships between concepts, with the strength of the relationships depending on the various contexts of the application domain. In philosophical terms, the process of inductive syllogism, as Aristotle calls the process of acquiring first principles from experience, cannot produce absolute knowledge. An inductively produced implication $X \to Y$ does not formally express the "necessity" of Y when X is known, but rather an argument for Y when X is known, thus making Y "probable" in this particular case, as the philosopher David Hume [63] suggests. Recent work seeks to acquire knowledge that is directly expressible in the form of such arguments [64].

VII. Towards a Future of Cognitive Systems

In its early days, Artificial Intelligence had sought to understand human intelligence and to endow machines with human-like cognitive abilities. Since then, however, AI has evolved primarily as an engineering discipline, placing emphasis on the development of useful specialized tools, and effectively abandoning the scientific inquiry into what constitutes intelligent behavior. In a sense, then, cognitive systems embody a modern realization of the need to return to AI's scientific roots, while adopting the engineering goal of developing useful systems.

This paper has sought to argue that computational argumentation in AI can offer a principled basis for the development of cognitive systems for everyday tasks. We have discussed work from Psychology showing that human cognition is inherently argumentative, and we have demonstrated that computational argumentation naturally encompasses several salient features of everyday human cognition — contra to the prevalent (even if implicit) assumption that classical logic can serve this role.

Given this new logical foundation for cognitive systems, one could reasonably ask whether it would necessitate a novel computing architecture on which to be realized, much like the Von Neumann architecture realizes classical (Boolean) logic. A neural-symbolic architecture (see, e.g., the workshop series "NeSy: Neural-Symbolic Learning and Reasoning") could potentially serve this role, with excitatory and inhibitory links implementing supports and attacks within an argumentation framework. Such an architecture could also allow the utilization of modern advances in deep learning, integrating within the reasoning architecture the process of learning and revision.

The view of logic reasoning as capturing the "laws of human thought" has served AI and Computer Science well. With an eye towards the development of cognitive systems, we would posit that it would be equally serving to view computational argumentation as capturing the "laws of common sense thought".

Acknowledgments

We would like to thank our colleagues, Irianna Diakidoy, Bob Kowalski, Hugo Mercier, Rob Miller, Pavlos Moraitis, Nikos Spanoudakis, Francesca Toni, and György Turán, for useful discussions and joint work on the topics of this paper.

References

[1] L. Michael, A. Kakas, R. Miller, and G. Turán, "Cognitive Programming," in *Proceedings of the 3rd International Workshop on Artificial Intelligence and Cognition*, 2015, pp. 3–18.

[2] P. Langley, "The Cognitive Systems Paradigm," *Advances in Cognitive Systems*, vol. 1, pp. 3–13, 2012.

[3] J. Evans, "Logic and Human Reasoning: An Assessment of the Deduction Paradigm," *Psychological Bulletin*, vol. 128, no. 6, pp. 978–996, 2002.

[4] D. Kahneman and A. Tversky, "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology*, vol. 3, no. 3, pp. 430–454, 1972.

[5] L. Rip, *The Psychology of Proof: Deductive Reasoning in Human Thinking*. MIT Press, 1994.

[6] P. Johnson-Laird, "Rules and Illusions: A Critical Study of Rips's The Psychology of Proof," *Minds and Machines*, vol. 7, no. 3, pp. 387–407, 1997.

[7] A. Kakas, F. Toni, and P. Mancarella, "Argumentation Logic," in *Proceedings of the 5th International Conference on Computational Models of Argument*, 2014, pp. 345–356.

[8] P. Cheng and K. Holyoak, "Pragmatic Reasoning Schemas," *Cognitive Psychology*, vol. 17, no. 4, pp. 391–416, 1985.

[9] G. Stoerring, "Experimentelle Untersuchungen ueber einfache Schlussprozesse," *Archiv fuer die gesammte Psychologie*, vol. 11, no. 1, pp. 1–127, 1908.

[10] P. Johnson-Laird, *Mental Models*. Cambridge University Press, 1983.

[11] P. Johnson-Laird and M. Steedman, "The Psychology of Syllogisms," *Cognitive Psychology*, vol. 10, no. 1, pp. 64–99, 1978.

[12] P. Johnson-Laird and R. Byrne, *Deduction*. Lawrence Erlbaum Associates, 1991.

[13] K. Stenning and M. van Lambalgen, *Human Reasoning and Cognitive Science*. MIT Press, 2008.

[14] ——, "Reasoning, Logic, and Psychology," *WIREs Cognitive Science*, vol. 2, no. 5, pp. 555–567, 2010.

[15] R. Kowalski, *Computational Logic and Human Thinking: How to Be Artificially Intelligent*. Cambridge University Press, 2011.

[16] H. Mercier and D. Sperber, "Why Do Humans Reason? Arguments for an Argumentative Theory," *Behavioral and Brain Sciences*, vol. 34, no. 2, pp. 57–74, 2011.

[17] S. Toumlin, *The Uses of Argument*. Cambridge University Press, 1958.

[18] C. Perelman and L. Olbrechts-Tyteca, *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, 1969.

[19] J. Pollock, "Defeasible Reasoning," *Cognitive Science*, vol. 11, no. 4, pp. 481–518, 1987.

[20] J. McCarthy, "Programs with Common Sense," in *Semantic Information Processing*. MIT Press, 1968, pp. 403–418.

[21] C. Alchourrón, P. Gärdenfors, and D. Makinson, "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions," *Symbolic Logic*, vol. 50, no. 2, pp. 510–530, 1985.

[22] N. Rescher, *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*. State University of New York Press, 1977.

[23] P. Valesio, *Novantiqua: Rhetorics as a Contemporary Theory*. Indiana University Press, 1980.

[24] A. Gardner, *An Artificial Intelligence Approach to Legal Reasoning*. MIT Press, 1987.

[25] T. Gordon, *The Pleadings Game: an Artificial Intelligence Model of Procedural Justice*. Kluwer Academic Publishers, 1995.

[26] R. Loui and J. Norman, "Rationales and Argument Moves," *Artificial Intelligence and Law*, vol. 3, no. 3, pp. 159–189, 1995.

[27] P. Dung, "On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games," *Artificial Intelligence*, vol. 77, no. 2, pp. 321–357, 1995.

[28] A. Kakas, R. Kowalski, and F. Toni, "Abductive Logic Programming," *Logic and Computation*, vol. 2, no. 6, pp. 719–770, 1992.

[29] A. Bondarenko, F. Toni, and R. Kowalski, "An Assumption-Based Framework for Non-Monotonic Reasoning," in *Proceedings of the 2nd International Workshop on Logic Programming and Non-Monotonic Reasoning*, 1993, pp. 171–189.

[30] T. Bench-Capon and P. Dunne, "Argumentation in Artificial Intelligence," *Artificial Intelligence*, vol. 171, no. 10–15, pp. 619–641, 2007.

[31] I. Rahwan and G. Simari, *Argumentation in Artificial Intelligence*. Springer Publishing Company, 2009.

[32] M. Lippi and P. Torroni, "Argumentation Mining: State of the Art and Emerging Trends," *ACM Transactions on Internet Technology*, vol. 16, no. 2, pp. 10:1–10:25, 2016.

[33] F. Toni and P. Torroni, "Bottom-Up Argumentation," in *Proceedings of the 1st International Workshop on Theories and Applications of Formal Argumentation*, 2012, pp. 249–262.

[34] T. Bench-Capon, "Persuasion in Practical Argument using Value-Based Argumentation Frameworks," *Logic and Computation*, vol. 13, no. 3, pp. 429–448, 2003.

[35] N. Karacapilidis and D. Papadias, "Computer Supported Argumentation and Collaborative Decision Making: The HERMES System," *Information Systems*, vol. 26, no. 4, pp. 259–277, 2001.

[36] L. Amgoud and H. Prade, "Using Arguments for Making and Explaining Decisions," *Artificial Intelligence*, vol. 173, no. 3–4, pp. 413–436, 2009.

[37] A. Kakas, L. Amgoud, G. Kern-Isberner, N. Maudet, and P. Moraitis, "ABA: Argumentation Based Agents," in *Proceedings of the 8th International Workshop on Argumentation in Multiagent Systems*, 2011, pp. 9–27.

[38] A. Kakas and P. Moraitis, "Argumentation Based Decision Making for Autonomous Agents," in *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, 2003, pp. 883–890.

[39] B. Verheij, "Formalizing Value-Guided Argumentation for Ethical Systems Design," *Artificial Intelligence and Law*, vol. 24, no. 4, pp. 387–407, 2016.

[40] N. Foo and Q. Vo, "Reasoning about Action: An Argumentation-Theoretic Approach," *Artificial Intelligence Research*, vol. 24, pp. 465–518, 2005.

[41] E. Hadjisoteriou and A. Kakas, "Reasoning about Actions and Change in Argumentation," *Argument and Computation*, vol. 6, no. 3, pp. 265–291, 2015.

[42] A. Kakas, R. Miller, and F. Toni, "An Argumentation Framework of Reasoning about Actions and Change," in *Proceedings of the 5th International Conference on Logic Programming and Non-Monotonic Reasoning*, 1999, pp. 78–91.

[43] L. Michael, "Story Understanding... Calculemus!" in *Proceedings of the 11th International Symposium on Logical Formalizations of Commonsense Reasoning*, 2013.

[44] I. Diakidoy, A. Kakas, L. Michael, and R. Miller, "Story Comprehension through Argumentation," in *Proceedings of the 5th International Conference on Computational Models of Argument*, 2014, pp. 31–42.

[45] ——, "STAR: A System of Argumentation for Story Comprehension and Beyond," in *Proceedings of the 12th International Symposium on Logical Formalizations of Commonsense Reasoning*, 2015, pp. 64–70.

[46] Y. Dimopoulos, L. Michael, and F. Athienitou, "Ceteris Paribus Preference Elicitation with Predictive Guarantees," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, 2009, pp. 1890–1895.

[47] L. Michael and E. Papageorgiou, "An Empirical Investigation of Ceteris Paribus Learnability," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2013, pp. 1537–1543.

[48] D. Lenat, "CYC: A Large-Scale Investment in Knowledge Infrastructure," *Communications of the ACM*, vol. 38, no. 11, pp. 33–38, 1995.

[49] F. Suchanek, G. Kasneci, and G. Weikum, "YAGO: A Core of Semantic Knowledge," in *Proceedings of the 16th International Conference on World Wide Web*, 2007, pp. 697–706.

[50] R. Speer and C. Havasi, "ConceptNet 5: A Large Semantic Network for Relational Knowledge," in *The People's Web Meets NLP: Collaboratively Constructed Language Resources*, I. Gurevych and J. Kim, Eds. Springer Berlin Heidelberg, 2013, pp. 161–176.

[51] C. Rodosthenous and L. Michael, "A Hybrid Approach to Commonsense Knowledge Acquisition," in *Proceedings of the 8th European Starting AI Researcher Symposium*, 2016, pp. 111–122.

[52] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. Hruschka Jr., and T. Mitchell, "Toward an Architecture for Never-Ending Language Learning," in *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 2010, pp. 1306–1313.

[53] L. Michael and L. Valiant, "A First Experimental Demonstration of Massive Knowledge Infusion," in *Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning*, 2008, pp. 378–389.

[54] L. Michael, "Reading Between the Lines," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, 2009, pp. 1525–1530.

[55] ——, "Machines with WebSense," in *Proceedings of the 11th International Symposium on Logical Formalizations of Commonsense Reasoning*, 2013.

[56] L. Valiant, "Robust Logics," *Artificial Intelligence*, vol. 117, no. 2, pp. 231–253, 2000.

[57] L. Michael, "Autodidactic Learning and Reasoning," Doctoral Dissertation, Harvard University, Cambridge, Massachusetts, U.S.A., 2008.

[58] ——, "Simultaneous Learning and Prediction," in *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning*, 2014, pp. 348–357.

[59] H. Skouteli and L. Michael, "Empirical Investigation of Learning-Based Imputation Policies," in *Proceedings of the 2nd Global Conference on Artificial Intelligence*, 2016, pp. 161–173.

[60] L. Michael, "Causal Learnability," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, 2011, pp. 1014–1020.

[61] ——, "The Disembodied Predictor Stance," *Pattern Recognition Letters*, vol. 64, no. C, pp. 21–29, 2015, Special issue on 'Philosophical Aspects of Pattern Recognition'.

[62] ——, "Introspective Forecasting," in *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 2015, pp. 3714–3720.

[63] D. Hume, *A Treatise of Human Nature*, L. Selby-Bigge, Ed. Clarendon Press, 1888, Originally published in 1738-1740.

[64] L. Michael, "Cognitive Reasoning and Learning Mechanisms," in *Proceedings of the 4th International Workshop on Artificial Intelligence and Cognition*, 2016.