A Computational Cognitive Model of Reasoning in Tibetan Buddhist Monastic Debate

Stefan Riegl (stefan.riegl@student.ru.nl)
Radboud University, Comeniuslaan 4
6525 HP, Nijmegen, The Netherlands

Marieke van Vugt (m.k.van.vugt@rug.nl)
University of Groningen, Nijenborgh 9
9747 AG Groningen, The Netherlands

Abstract

In Tibetan monasteries, the education system relies heavily on a very specific style of debating that is at once exhilarating and intellectually rigorous. Relatively little research has been done on the psychological and neural mechanisms of this debate, which may be an interesting method for education around the world. Hence the formation of a theory of this practice is important. Here we present a computational theory of Tibetan monastic debate implemented in the ACT-R cognitive architecture. We complement the ACT-R model with graph theory to represent knowledge and show how we can capture the dynamic flow of a debate in our model. Future research should validate the model in its native population and enrich it with more detailed strategies. Nevertheless, we think it provides an interesting example of how the interactive process of debating can be modelled.

Keywords: Reasoning; monastic debate; ACT-R; reasoning agents; graph-theory; logic; knowledge representation.

Introduction

Different cultures have different education methods. While Western education has been well studied, other forms of teaching and learning exist. One intriguing method is Tibetan monastic debate, a part of analytical meditation practises; a method that is practised in a dyad and is said to date back to the 10th century (van Vugt et al., 2019). It bears similarities to the Socratic method.

This method of debating is characterised by vigorous exercise, excitement, as well as focus on critical thinking and examining many different perspectives. In other words, a dialectic method like the monastic debate would introduce an inherently social aspect to studying, that would complement more individual-centred study approaches. With its game-like structure and active methods, debate may activate students and help them to think more critically.

Repeated practise of debate is likely to have substantial impact on the cognitive development of the debater. Recent research (van Vugt et al., 2020; van Vugt et al., 2019) speculated that debate has a positive effect on executive functions, such as e.g. critical thinking, emotion regulation, or social cognition. However, up until today no evidence has been found that debate trains cognitive functions. Further research is necessary to identify the cognitive skills that are at play during monastic debate.

Modelling debate helps drawing up specific hypotheses what cognitive skills are required for debate. With a model predictions can be made that certain cognitive skills could be improved after intense practice (Taatgen, 2013) and what can happen to debate if a certain skill is missing or not available due to e.g. exhaustion or speed pressure.

Monastic debate

Monastic debate is part of all Tibetan monastic traditions, but with several hours a day over the course of 20 years (Dreyfus, 2003) it is most intensively practised in the Geluk school. In general, monastic education is centred around the study and memorisation of Buddhist scriptures. Debate is used to test and deepen the debater's knowledge, and to sharpen skills in critical thinking and logical reasoning. The general method of debate can be understood as a form of *reductio ad absurdum* common in logical argumentation. What makes this debate form unique is that it aims at uncovering shortcomings and inconsistencies in in the debater's knowledge and understanding. This sets Monastic debate apart from other common debating styles, where debaters attempt to defeat the other debater with stronger arguments. An example of monastic debate can be seen in figure 1.

Debate is a dialogue between a “challenger” and a “defender” (Perdue, 2014) and typically covers topics of recently discussed lessons in Buddhist philosophy. Generally the challenger proposes statements to which the defender responds. The defender has to choose between accepting and rejecting a statement, while ensuring that no statement is accepted that contradicts an earlier one. Statements are intended to probe the consistency of a particular philosophical position of the other debater. However, they are not required to be rational or correct, as they can be understood as a tool to explore the consequences of adopting a particular philosophical position. Monastic debate follows a formal schema that includes choreographic elements like shouting and clapping, but also statement-response patterns, as outlined in table 1.

The structure of debate just discussed may lend it well as a tool for scientists. According to monks from the Sera Jey monastery, debate supports the investigation of a topic from various perspectives by exploring the consequences of adopting a particular position. So on the one hand, investigating the research hypotheses by means of debate can help to reveal latent, inconsistent or overly restricting assumptions, which can then be resolved. On the other hand debate is an interaction form that can promote novel ways of thinking about a topic and as such can lead to fresh insights into e.g. scientific results. Moreover, excellent ownership of the material is required to keep up with the speed of debate and maintain its formal structure, which pushes the boundaries of scientists in an engaging way.

Previous research

Scientific research on the psychological and neural mechanisms of debate is very limited. The monastic curriculum (Liberman, 2007; Dreyfus, 2003) and debate as a whole (Perdue, 2014) were discussed. It was suggested that debate might share commonalities with the Socratic method or might be a "mode of inquiry" (Dreyfus, 2003). However there seems to be a paucity of research on what effects debate has on the individual, especially from a cognitive science perspective.

The single, empirical study that has so far been done (van Vugt et al., 2020) showed that mid-frontal theta oscillations (neural correlate for concentration) increased during debate. It also showed that frontal alpha oscillations synchronised between debaters when they agreed compared to when they disagreed. The researchers theorised that successful debate requires a rich set of cognitive skills and that repeated exercise over years fosters those skills. A conceptual model was proposed (van Vugt et al., 2019) that suggests how debate requires skills in focused attention, working memory, and logical reasoning, but also emotion regulation and mental flexibility. Further research could provide more evidence what particular cognitive processes are relevant and change during training in this method.

C: Dhih! The subject, in just the way *[Manjushri debated]*. Is whatever is a colour necessarily red?
D: I accept *[that whatever is a colour is necessarily red]*.
C: It follows that whatever is a colour is necessarily red.
D: I accept it.
C: It *[absurdly]* follows that the subject, the colour of a white religious conch, is red.
D: Why *[is the colour a white religious conch red]*?
C: Because of being a colour. You asserted the pervasion *[that whatever is a colour is necessarily red]*.
D: The reason *[that the colour of a white religious conch is a colour]* is not established.
C: It follows that the subject, the colour of a white religious conch, is a colour because of being white.
D: The reason *[that the colour of a white religious conch is white]* is not established.
C: It follows that the subject, the colour of a white religious conch, is white because of being one with the colour of a white religious conch.
D: I accept that the colour of a white religious conch is white.

Figure 1: An example of Tibetan Buddhist monastic debate, adapted from (Perdue, 2014). *C* denotes the challenger and *D* the defender.

Statement type, example (challenger)	**Possible responses (defender)**
Two-part debate "Red is a visual form"	• "I accept." • "Why?" / "No."
Three-part debate "Red is a visual form, because it is a colour."	• "I accept." • "Reason not established." • "No pervasion."
Inquire about reason "Red is a visual form, because. . . "	• ". . . because red is a colour."
Request an example (context-dependent) "Posit it!"	[Context: Something that is a bird and that cannot fly.] • "A penguin."

Table 1: Typical statements types of the challenger and possible responses of the defender.

A relevant starting point for modeling debate are models of logical reasoning and human interaction. Several of these models have been implemented in the ACT-R cognitive architecture (Anderson et al., 2004). Analogical reasoning for example was modelled to solve Raven's Progressive Matrices (Ragni & Neubert, 2012), which are frequently used in IQ-tests. Objects in the cells of matrices are decomposed into different attributes, that are used to identify the rules that allow predicting the missing element of the matrix. Inference rules are implemented as ACT-R productions and the rule currently being checked is encoded in a chunk slot. Similar mechanisms could be used in a model of monastic debate.

In a different study (Ghosh, Halder, Sharma, & Verbrugge, 2015) strategies based on forward and backward induction in sequential games were investigated. A logic language was created to describe strategies and beliefs, which was then used heavily to model reasoning rules in a cognitive model. Strategies were selected based on expected payoffs. This mechanism would be more challenging to implement for monastic debate, since payoffs cannot be easily defined given that they are defined by high-level attributes such as novel insights.

A crucial ingredient for debating is theory of mind. There exist several computational models of theory of mind, which describe theory of mind as a sequence of reasoning steps with complexity dependent on the order of theory of mind, e.g. (Meijering, Taatgen, van Rijn, & Verbrugge, 2014). Successful debating is likely to involve theory of mind as well in the sense that possible moves of the opponent must be predicted to effectively trap them into a contradiction. Several strategies employed a theory of mind, including second-level theory of mind and higher. Another strategy was chosen, if the current one proved unsuccessful after a number of iterations.

A debate model necessarily requires interaction. A basis for modelling this can be found in modelling negotiation skills (Stevens et al., 2018), where a cognitive model was created capable of using different strategies. The model made decisions based on instances of the current state of the game using ACT-R's partial matching feature. Instances were either encoded into the model's declarative memory or learned (instance-based learning). Humans who played against the cognitive model then showed improvement in their negotiation skills in terms of payoff in the Game of Nines. A simple graphical computer interface allowed interaction with the model. For a model of debate this shows that is well possible to allow interaction to train improve task-specific skills, and to let a cognitive model pursue and switch between strategies, which is important for the challenger on a experienced level.

To conclude, models of debate can be based on previous models of sequential games, logical inference, and theory of mind. On the basis of this insight we attempted to model monastic debate.

Methods

ACT-R

The cognitive model of debate was implemented in ACT-R, after previous research showed that it is well possible to model interacting and reasoning agents with this cognitive architecture (Ragni & Neubert, 2012; Ghosh et al., 2015; Meijering et al., 2014; Stevens et al., 2018). ACT-R is an excellent tool for this, because it is commonly used, well maintained, and several relevant models have been created in the past of which parts can form the basis for modelling monastic debate.

ACT-R largely consists of modules that represent executive functions of humans, e.g. visual perception or motor execution. The behaviour of the model is defined by a production system, which represents procedural memory. A single production rule can be understood as versatile if-then-statement. Model behaviour can be modulated by so-called subsymbolic features. For example, based on spreading activation or the similarity of chunks in the declarative memory, it is possible to model phenomena such as reasoning errors caused by confusing concepts or by mistaking relations between concepts.

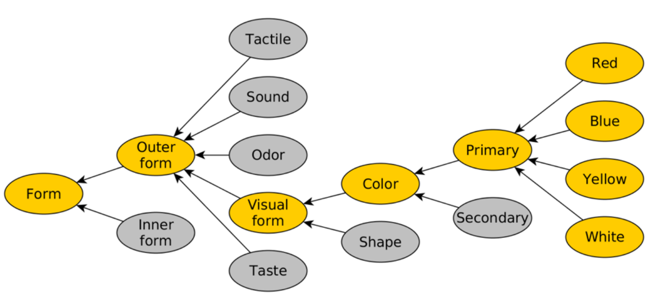

Figure 2: A sub-graph of the "colour map" (Tharpa & Tsultrim, 2012), a hierarchically organised topic of debate often used by learning debaters. Child nodes of the nodes marked in grey have been omitted for brevity.

Knowledge representation

Reasoning agents need something to reason about. Traditionally, the content of a debate is a topic from Buddhist teachings which was discussed in recent classes of the monastic curriculum. Such material is the first choice as object of debate for the model, because such traditional topics were used and explicated in real debate over thousands of years. Moreover, using the traditional debate ontology allows us to validate the model with Tibetan monks more easily (see the section on validation below).

In their first classes learning debaters usually reason about ontologies of colours and forms (Perdue, 2014), mental and physical phenomena, causation, and other topics. Each debate topic is limited to a finite set of concepts and how those concepts relate to each other, for example in *is-a*, *has-a*, *causes*, and other relations. For the cognitive model the topic of colours and forms is used as the conceptual space, which is based on *is-a* relations between concepts. To this end the knowledge tree found in the literature was replicated as knowledge base for the model.[1] An excerpt of the knowledge tree can be seen in figure 2.

Most of the listed topics found in debate tutorials that were prepared for a Western audience (Tharpa & Tsultrim, 2012) showed a hierarchical structure.[2] Hierarchically structured data has the advantage that it can be described as a tree in our model by means of graph theory. Graph theoretical approaches have been studied and applied extensively in the past, which is know-how that can be drawn from. Using graph theory as basis to represent knowledge is helpful and simplifies the conception and description of ways to manipulate a knowledge base.

While the data found in Buddhist literature can be expressed as a tree, for our formal description it is only assumed that a debater's knowledge can be represented by a directed graph.[3] For a debater d (challenger or defender) in one instance of a debate, let $G_d = (V_d, E_d, P)$ be a directed graph representing the debate-specific knowledge of d, where V_d are the nodes, E_d are the edges of the graph, and P is a single binary predicate. An example of a knowledge graph including nodes and edges can be seen in figure 2.

More specifically, $V_d = \{v_1, v_2, \ldots, v_n\}$ is a set of nodes that represents the concepts that d knows. For example, v_1 might represent the concept "colour red", v_2 the concept "primary colour", and v_3 the general concept "colour". P is a binary predicate that represents a transitive relation between concepts, for example an *is-a* relation. Transitive relation here means that if $P(v_1, v_2)$ and $P(v_2, v_3)$, then $P(v_1, v_3)$ for

[1] While learning the correct relation between concepts is a desired practice and learning new concepts in conversational agents is surely possible in principle, it is not clear whether learning new concepts is common in real debate.

[2] The reviewed literature covers topics of the first year of the monastic curriculum at the Sera Jey monastery. We do not know yet whether topics from higher years are also organised hierarchically.

[3] To perfectly represent traditional material structured in one tree, a graph needs to be acyclic and only have a single parent node.

$\{v_1, v_2, v_3\} \in V_d$. For each directed edge $(v_1, v_2) \in E_d$ (with $E_d \subseteq V_d \times V_d$) it holds that $P(v_1, v_2)$. For example, if P represents an *is-a* relation, v_1 represents the concept "colour red", v_2 represents the concept "primary colour" and $(v_1, v_2) \in E_d$ is a directed edge, then d can infer that "colour red" is a "primary colour".

In the knowledge graph concepts are represented by nodes, while in the model they are represented by ACT-R chunks and only include a slot for the name. A concept is retrieved based on its name, which may appear as a word when processing auditory input. Pervasions (directed edges in the graph) are modelled by chunks with two slots that contain those concepts. In ACT-R this is expressed as a chunk as follows in the example of the pervasion "color is a visual form":

```
(per_color_is_a_visual_form
 isa          pervasion
 pervaded     c_color
 pervader     c_visual_form)
```

The pervaded and the pervader can be likened to the antecedent and consequent in a material implication in formal logic. Depending on the specific debate statement, the model recalls a pervasion by matching either one or both concepts, as described in the next section. If the model encounters an unknown or otherwise unexpected word, the model responds that it cannot understand the input given.

Reasoning with graphs

Knowing details does not mean one knows the bigger picture. If a debater d has knowledge in form of a tree $G_d = (V_d, E_d, P)$ and d knows $e_1 = (v_a, v_b)$ and $e_2 = (v_b, v_c)$ (i.e. $\{e_1, e_2\} \subseteq V_d$), then d however does not necessarily know $e_3 = (v_a, v_c)$ (i.e. not generally $e_3 \in V_d$). However, d may deduce this new fact based on the property of transitivity of P. This then adds a new edge to V_d (i.e. $V_d' = V_d \cup \{e_3\}$). In other words, if P represents an *is-a* relation, d can apply the syllogism "All [tones of] red are primary colours; All primary colours are colours; Therefore, all [tones of] red are colours". ACT-R was used to create a model of both the defender and the challenger each that implements this form of deductive reasoning. At the moment only the defender uses this inference to gain new knowledge.

One of the statements the challenger might issue is the "three-part debate" statement. For example, if d is a challenger and P represents an *is-a* relation between concepts, then three-part debate typically follows the form "Take the subject $\langle subject \rangle$, it follows it is $\langle predicate \rangle$, because it is $\langle reason \rangle$." where the three placeholders represent concepts in V_d. When a three-part debate statement is given to the defender, deductive inference allows to to confirm or reject the statement. For example, let the defender know "Socrates is a man" and "All men are mortal". If the challenger inquires "Is Socrates a man?" the defender can immediately respond with "I accept.", since the defender possesses exactly that requested piece of knowledge. However, if the challenger asks "Is Socrates mortal?", then the defender should be able to infer the proposition in question ("Socrates is mortal.") from given knowledge.

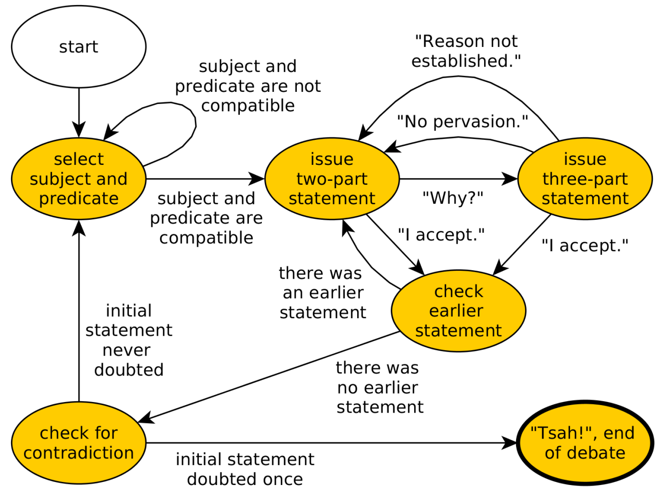

Figure 3: A visualisation of the course of debate with limited ways to respond, from the point of view of a challenger.

Based on the *negation as failure*-principle, the defender model will accept the logical proposition of the debate statement, if the proposition can be inferred deductively in one or more steps based on the knowledge the defender has. If the logical proposition cannot be inferred, it will be rejected with the words "Reason not established" or "No pervasion" (just like in real debate), depending on at which point the inference fails.

Model evaluation

Commonly ACT-R models are evaluated by comparing model and human performance with regards to reaction times, accuracy of task performance, or fMRI or EEG measurements. Refining the model until its performance matches human performance, including error during performance, is a method to validate the model.

However, monastic debate does not follow a typical task paradigm, which makes it hard to apply standard methods. Due to the complexity of debate it is difficult to measure reaction times. Capturing the accuracy of a debater is difficult, as there are no typical metrics to measure whether a debater performs well during debate. Debate measures related to accuracy that are typical for experimental tasks might be the number and type of reasoning errors, the length and duration of debate, how often the defender responds in a certain way, or how often the challenger changes strategies. However, those data first need to be collected in some way and even then comparison might still be tricky, since any two debates rarely follow the same course.

To acquire data from the model two interfaces were created for interaction with the model. A terminal-based interface allows for quick testing and automation during model creation. The other interface is runs in web browsers and offers a few conveniences. Graphical user interfaces are often more accessible, can provide an easier way to change relevant parameters, and generally allow a better user experience for a less technology-savvy audience. This is important, as the browser-based interface is intended to be used for data acquisition and model validation in the future. A welcome side effect in the spirit of open science is that the browser-based design is easy to adapt for access via the internet[4].

Results

While the defender remains mostly reactive, the challenger needs to plan ahead what statements to issue at what time, to eventually lead the defender to accept a contradicting statement. For a better understanding of the directions a debate can take, it might help to focus not on one of the debaters specifically, but their interactions over time. The possible paths of a very limited form of debate are visualised in figure 3. The diagram can be read like a state diagram of a finite state automaton. Note that the diagram contains the essential two-part and three-part debate statements, but no other types of statements. To capture the more diverse and complex human debates fully the diagram would need to be extended.

An example interaction via the terminal-based interface between a human defender and the challenger model can be seen in figure 4. In the example, the relation between concepts is an *is-a* relation, so the two-part and three-part debate statements take the general form "Take the subject $\langle subject \rangle$, it follows it is $\langle predicate \rangle$[, because it is $\langle reason \rangle$].".

In the beginning the challenger posits a two-part debate statements until the the defender does not accept. This provides the starting point for the debate and the challenger now tries to make the defender accept the same statement. As next step the two-part debate statement will be extended to a three-part debate statement by providing a $\langle reason \rangle$. The $\langle reason \rangle$ essentially is one of the concepts represented by a node in the knowledge graph, that lies on the path between the nodes for the $\langle subject \rangle$ and the $\langle predicate \rangle$ in the two-part debate statement. If the defender disagrees that "$\langle subject \rangle$ is a $\langle reason \rangle$", a "Reason not established" response follows, but if the defender disagrees that "$\langle reason \rangle$ is a $\langle predicate \rangle$", a "No pervasion" response follows. In both cases the challenger then knows the two concepts for which an *is-a* relation is denied and so issues a new two-part debate statement based on those two concepts[5]. After issuing this new statement the same procedure as just described follows, so this can be considered a new step in a nested or recursive procedure.

However, the defender might reject a two-part debate statement, but not the respective three-part debate statement. This is the case, when the defender has the relations "$\langle subject \rangle$ is a $\langle reason \rangle$" and "$\langle reason \rangle$ is a $\langle predicate \rangle$" represented as edges in the knowledge graph. Then the defender accepts the three-part statement, which could be considered going back one step in the recursive procedure. If the defender accepts the statement that was rejected in the beginning, the challenger may declare the uncovering of a contradiction with a "Tsah!" response and the debate ends.

In contrast to the challenger model, the defender model is simpler, yet also based on the course of debate shown in figure 3. Generally, a two-part debate statement is accepted, if the defender knows that "$\langle subject \rangle$ is a $\langle predicate \rangle$". For a three-part debate statement the defender will check whether "$\langle subject \rangle$ is a $\langle reason \rangle$" and "$\langle reason \rangle$ is a $\langle predicate \rangle$" and react as described above. Accepting a three-part debate statement means that the model "has an insight" and learned that "$\langle subject \rangle$ is a $\langle predicate \rangle$", i.e. added an edge to the knowledge graph.

This shows that the model can function in the simplified world of the first-year debate material. The model can take the role of both a challenger and a defender that evaluate two-part and three-part debate statements in the current version of the model. Reasoning is performed error-free on the basis of complete (challenger) or incomplete (defender) knowledge.

C:	Take the subject rough, it follows that it is form.
D:	`why`
C:	Take the subject rough, it follows that it is form, because it is outer form.
D:	`no-reason`
C:	Take the subject rough, it follows that it is outer form.
D:	`why`
C:	Take the subject rough, it follows that it is outer form, because it is tactile.
D:	`no-reason`
C:	Take the subject rough, it follows that it is tactile.
D:	`why`
C:	Take the subject rough, it follows that it is tactile, because it is derived from element.
D:	`accept`
C:	Take the subject rough, it follows that it is outer form, because it is tactile.
D:	`accept`
C:	Take the subject rough, it follows that it is form, because it is outer form.
D:	`accept`
C:	Tsah!

Figure 4: The model in action: An example debate with a challenger model (denoted *C*) and a human defender (denoted *D*). The simplified responses of the human defender are interpreted as the written out, formally correct debate statements.

[4]At the time of writing, a recent state of the model can be found at `https://blueparrot.pythonanywhere.com/`. This address might updated in the future.

[5]From an algorithmic point of view it is not necessary to restate the previously rejected part of the three-part as two-part debate statement, as the defender will always consistently respond with "Why?" to it. However, in real debate it is generally considered good practice to be explicit, which should then also be captured by a model.

Discussion

Validation

Monastic debate involves many cognitive processes and different ways to interact with the environment, including high-level executive functions such as theory of mind. Due to the complexity of debate it seems difficult to apply validation methods that are typically used for ACT-R models, such as correlating reaction times. In addition, it appears tricky to test isolated parts of debate, as essential aspects of this integrative practise might simply get lost in a controlled lab setting. Investigating cognition "in the wild" is an objective of the scientific community of macrocognition and their methodology such as Cognitive Task Analysis (Crandall, Klein, Klein, & Hoffman, 2006) might help to investigate debate.

However, certain quantitative properties of debate could be collected and compared, such as the level of a debater's expertise, the number of agreements between debaters (van Vugt et al., 2019), or the number of reasoning errors per debate. As an additional way to collect such data and as a separate form of model validation a "Monastic Turing Test" could be performed, i.e. a Turing Test[6] that allows a human judge to rate responses from another debater. The human judge is debating with the partner via a text chat and normalised responses, but the judge does not know whether the other debater is another human or one of the challenger or defender models.

To be able to perform such a monastic Turing test, a browser-based interface was created to interact with the model. Such an interface has the additional benefit that researchers can collect normalised data in an automated way, which is helpful for further assessment of debates and comparison with model behaviour.

Before starting a debate via the interface, a human debater is able to select whether the role of the challenger should be filled by a human or by the cognitive model, and similarly for the defender. Random assignment of the judge to a model or a human is possible, which allows to realise the Monastic Turing Test. If a certain model is judged to be more human than other models after a number of iterations, then the model may follow certain strategies or make certain errors that resemble those of humans more closely.

Contrasting candidate models that differ in ACT-R model parameters or implemented strategies can be considered a "relative comparison", which does not yet validate a model (Palminteri, Wyart, & Koechlin, 2017). Hence it is imperative to evaluate single models based on their "generative performance", i.e. whether and how well a model is able to reproduce evidence, including different debate strategies. Nevertheless, we think that our model does currently not match human performance well, because it does not make errors yet.

[6] In the text the Turing Test is based on the modern, most common interpretation, not in the sense of the "imitation game" as originally proposed by Allan Turing.

Reasoning errors

Errare humanum est. Humans do not always perform optimal and one way to make the model behave more human-like is to introduce errors. In human debates however mistakes do occur frequently, especially due to memory failures or time pressure, since taking too long to respond is considered to be a weakness (Dreyfus, 2003). Two common sources of errors appeared to be an inconsistent body of knowledge and an incorrect application of rules of logical inference for debate (reasoning errors).

Errors due to an inconsistent knowledge were modelled by a manipulated knowledge graph, where edges were added or removed such that the formal definition of a knowledge graph is violated. In those cases the model then returned erroneous responses, for example the model rejected that red is a primary colour or accepted that primary colours are both colours and sounds.

A reasoning error is given, when only valid premises are given (drawn from the body of knowledge or from the ongoing debate), but the reasoner still arrives at an incorrect conclusion. In human debate this is not uncommon, especially among novice debaters. ACT-R has a set of so called subsymbolic features like spreading activation or partial matching, which will aid the implementation of such reasoning errors.

There can be many reasons for reasoning errors, for example the defender can get nervous, confused by long statements, distracted by the environment, exhausted from a long debate or simply forget earlier statements. In ACT-R a defender's confusion of concepts could be modelled, for example, by encoding the similarity between (memory) chunks and allowing the retrieval of a similar, but incorrect chunk, when listening to the words of the challenger. Integrating reasoning errors into the model however is part of future research.

Future work

While considerable progress was made in creating a computational model of debate, there are limitations to the model.

Firstly, it is possible that our assumption of hierarchically-structured knowledge is too strong to represent the majority of real-life debates. Future models should attempt to relax this restriction. In addition, dealing with more general knowledge structures opens up the possibility to reason about more complicated topics of the monastic curriculum. Topics foreign to the Tibetan tradition might then also be considered, such as study material frequently discussed in Western education systems.

As mentioned in the earlier many forms of response types exist in debate, but they are used in different frequencies and not all are equally easy to formalise. Not all debates have a challenger who requests an example or the reason for a certain statement (see table 1). However the two-part and three-part debate statements and their responses are essential and could not be taken away without losing defining aspects of debate. Extending the model by adding more ways to interact makes the debate more engaging and the model more realistic.

Experienced challengers often switch between different debate styles, which might be less defined by what kinds of statements are issued, but more how the parts are combined to lead the debate. Such styles can differ in e.g. the use of analogies, the ratio between exploration and exploitation of debate subjects, referring back to previous debate subjects, the trade-off between fast vs. accurate responses, or attempts to trick the defender. Capturing such notions formally and integrating them into the model allows the model to match human behaviour more closely.

Conclusion

In this work an ACT-R model was created that uses graph theory for flexible and extensible knowledge representation. This innovative approach captures essential parts of simple debate instances, but also a cognitive process, which both have not been described formally before.

More generally, we think this research provides a promising early stage model of an interaction in a complex real-world environment. We are confident that by looking at tasks outside the ordinary domain, we can gain insights into cognitive skills that generalise better across all human beings, not only minds trained in Western education (Henrich, Heine, & Norenzayan, 2010).

Acknowledgements

The authors wish to thank Losang Donyo, a monk from the Sera monastery, for the joint sessions of debate and advisory support; and Amir Moye, who provided valuable input and feedback during model creation.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036.

Crandall, B., Klein, G., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Mit Press.

Dreyfus, G. (2003). *The sound of two hands clapping: The education of a tibetan buddhist monk*. Univ of California Press.

Ghosh, S., Halder, T., Sharma, K., & Verbrugge, R. (2015). Human strategic reasoning in dynamic games: Experiments, logics, cognitive models. In *International workshop on logic, rationality and interaction* (pp. 116–128).

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not weird. *Nature*, *466*(7302), 29–29.

Liberman, K. (2007). *Dialectical practice in tibetan philosophical culture: An ethnomethodological inquiry into formal reasoning*. Rowman & Littlefield Publishers.

Meijering, B., Taatgen, N. A., van Rijn, H., & Verbrugge, R. (2014). Modeling inference of mental states: As simple as possible, as complex as necessary. *Interaction Studies*, *15*(3), 455–477.

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in cognitive sciences*, *21*(6), 425–433.

Perdue, D. E. (2014). *The course in buddhist reasoning and debate: An asian approach to analytical thinking drawn from indian and tibetan sources*. Shambhala Publications.

Ragni, M., & Neubert, S. (2012). Solving raven's iq-tests: an ai and cognitive modeling approach. In *Proceedings of the 20th european conference on artificial intelligence* (pp. 666–671).

Stevens, C. A., Daamen, J., Gaudrain, E., Renkema, T., Top, J. D., Cnossen, F., & Taatgen, N. A. (2018). Using cognitive agents to train negotiation skills. *Frontiers in psychology*, *9*, 154.

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological review*, *120*(3), 439.

Tharpa, T., & Tsultrim, N. S. (2012). *Debate Primer Text 1 - A dialectic debate primer and method text for Tibetan monastic debate in the Tibetan language.* Liberation Publications. Retrieved 2020-04-03, from `http://tenzintharpa.com/download-library/` (3rd ed. 2012)

van Vugt, M. K., Pollock, J., Johnson, B., Gyatso, K., Norbu, N., Lodroe, T., ... Khechok, J. e. a. (2020). Inter-brain synchronization in the practice of tibetan monastic debate. *Mindfulness*, 1–15.

van Vugt, M. K., Moye, A., Pollock, J., Johnson, B., Bonn-Miller, M. O., Gyatso, K., ... others (2019). Tibetan buddhist monastic debate: psychological and neuroscientific analysis of a reasoning-based analytical meditation practice. In *Progress in brain research* (Vol. 244, pp. 233–253). Elsevier.