Models of Reasoning in Ancient China

Fenrong Liu, Jeremy Seligman, Johan van Benthem

1   Introduction

Imagine for a moment that you are given the opportunity of a trip to the ancient state of Qi (齊) in a suitably reliable time machine. Your destination is the famous Jixia (稷下) academy, which was founded during the reign of the King Wei of Qi (齊威王, 356–321 B.C.). When you arrive, the halls will be full of scholars debating such topics as the importance of ritual lǐ (禮), how to determine what is right yì (義), the relationship between names míng (名) and objects/reality shí(實), and of course, human nature xìng (性). Equipped with the tools of modern logic, you set out to make sense of these debates in the best way you can. One approach would be to start a school of logic, teaching the predicate calculus and set theory, in the hope that once the light of 21st century reason has been shed on the dark corners of the hall, many mysteries and sources of confusion would just evaporate. This may not be the best strategy.

One immediate problem is that these tools were developed in the 19th and 20th centuries to deal with problems in the philosophy of mathematics. For that, they are very well suited. Of course, they have gone on to find application in a much broader arena. But first and foremost, they are tools for reasoning about mathematical objects, timeless and discrete numbers, whose properties are more-or-less determinate, and about which the largest mystery concerns the treatment of infinity. No matter how much faith one has in the power of logic, the difficulty of showing the application of these techniques to the ancient debates must be appreciated.

Our approach will be different. To be sure, an understanding of our techniques depends on the same educational background as other parts of logic. But we will try to model the subject matter of the ancient debates directly, using only what is

needed when it is needed.[1] The aim of this paper is exploratory. We will propose a few models of the central concepts of classical Chinese philosophy. In particular, we analyze ideas of the School of Confucian and of the Mohist School, with the modest aim of indicating, in rough terms, how the techniques of modern logic[2] may be applied to matters of ancient concern. We do not claim that any of the ideas presented here give a correct or faithful account of their subject matter. They are models, that is all. Some may ultimately prove useful; others will have to be discarded along the way. We hope only to demonstrate the kind of application of modern techniques that we believe to be generally useful.

2 Yì: the right response

Chris Fraser, following ideas from Angus Graham, has proposed that we interpret ancient Chinese view of agency in terms of discrimination and response.[3] This is a very appealing idea. Our starting point will be to model this mathematically, in a very simple way, and see what can be done with it.

So, let us suppose that an agent encounters various *situations* and discriminates between them by performing certain *actions*. The responses of our agent can be modelled by a function r from situations to actions, so that in situation s, the agent responds with $r(s)$. We let S be the set of situations in which the agent responds.

Now, listening to the Jìxìa scholars, we gather that there is an important difference between responses. Some are *yì* (right/appropriate) and some are not. We will model the concept *yì* in a simple-minded way as a set Y of situation-response pairs. For $\langle s, r \rangle$ to be in the set Y is for the response r in situation s to be *yì*. Now we take a further step, trying to stay close to the action orientation of classical Chinese texts.We will take a situation to be determined by the responses it contains, and in fact, identify it with a set of responses. This requires a fairly liberal understanding of 'response', according to which everything participating in a situation (animate or inanimate) is understood to be 'responding' to something, if only in

---

[1]We do not deny the possibility of a translation or interpretation of our models back into something more standard: in fact, this is usually possible. But such a reduction would not add anything of importance.

[2]By 'modern logic' we do not mean the predicate calculus or set theory or any one symbolic system, classical or non-classical. Instead, we draw on the spirit of research in modern logic, especially those parts of logic commonly termed 'philosophical' or 'applied'. That is a spirit of 'anything goes,' in which problems and conundrums are approached on their own terms, without ideological bias towards any one system or set of techniques.

[3]See C.Fraser, Action and Agency in Early Chinese Thought, *Journal of Chinese Philosophy and Culture* 5: 217–239, 2009.

some minimal sense. We will now frame this as a definition.

> **Response structure** A *response structure* $\langle R, S, Y \rangle$ consists of a set R of *responses*, a set S of subsets of R (called *situations*) and a distinguished set Y of pairs $\langle s, r \rangle$ such that $r \in s$. Alternatively, the reader can thinking of Y as assigning a subset of 'correct responses' to each situation.

Simple as it is, this model raises many questions, which can lead to refinements of our interpretation of the model or point to areas of philosophical concern. Firstly, there is an obvious order on situations given by $s_1 \subseteq s_2$. What does this mean? At the extreme ends, does it make sense to talk of a maximal situation, consisting of all the responses in R, or a minimal situation $\emptyset$ containing no responses at all? In considering how to answer such questions, we are forced to be clearer about how to interpret 'response'. By modelling actual situations as sets of responses, we are committed to the impossibility of two different situations containing exactly the same responses. [4] There is of course a sense in which one can respond the same way in two different situations and, presumably, that all the participants in the situation can do this. But if there are two situations rather than one, then our model requires that the constituent responses also differ. Here it is important to keep in mind that our responses are concrete actions, not types. Two pulls of the bow string by Houyi (后羿) produce two perfect and so perfectly similar flights of an arrow, but they are still two different situations.

We will not continue down this metaphysical path, which already shows signs of introducing a discussion of the concept 'way' *dào* (道) that lies at the centre of all of Chinese philosophy. Instead, we will bring it back to the ethical. Modelling draws our attention to extreme cases. For example, it is natural to ask whether $\langle s, r \rangle$ belonging to Y depends on both s and r, or if not, whether there is a universal response r for which $\langle s, r \rangle \in Y$ for all s containing r. Such a response would be inherently *yì*, independent of the situation. Given the tendency of Chinese philosophers to emphasis the context dependency of central concepts, one might think that the answer is that there are many responses r such that for some situation s_1, $\langle s_1, r \rangle \in Y$ but for some other situation s_2, $\langle s_2, r \rangle \notin Y$. Here, however, we need to pay careful attention to the identity conditions for situations and responses. The appropriateness of a ritually prescribed action, such as bowing, depends on certain aspects of the situation, especially the relative status of the participants. Bowing may be appropriate in some cases, when in the presence of those with whom one has a certain social relationship, but not in others. But this

---

[4] In logicians' terms, we have adopted "response extensionality" for situations.

is *not* the kinds of context dependence that is at issue. The above point about concreteness of responses returns. 'Bowing' is a type of response, not a specific response, and our question concerns the appropriateness of a specific response, and whether this somehow varies according to the situation that it is deemed part of. [5]

3   Rituals and the correcting of names

The possibility of changing our responses (for the better) was of great interest to the scholars of ancient China. There were many ideas about how to achieve this and how not to. What characterizes the Confucian ways was that we should at least be guided by certain traditional forms of behaviour known as the *lǐ* (ritual), although different Confucians had different ideas about how this was to be put in practice. At first, the introduction of a new norm seems redundant. Why isn't it enough to aim for what is *yì* without having to worry about whether it also in accordance with the *lǐ*? Part of an answer to this question is clearly epistemological. Whereas it can be very difficult to know what is right in a given situation, the knowledge of what is proscribed or prohibited by ritual is at least humanly attainable.[6]

One idea of great significance for the study of logic in ancient China is that an understanding of *lǐ* consists primarily in an understanding of the correct use of language.This is the doctrine of 'correcting names' *zhèng míng* (正名), which is mentioned in the Analects of Confucius but developed much more fully in Xunzi. Here is what is said about *zhèng míng* in the Analects:

> Zilu said, 'The ruler of Wei awaits your taking on administration. What would be Master's priority?' The Master replied, 'Certainly - rectifying names!' ... If names are not rectified then language will not flow. If language does not flow, then affairs cannot be completed. If

---

[5]Some evidence for an acknowledgement of this stronger kind of dependence on situation comes from *Zhuangzi* (Chapter 20) Out hunting, he sees a cicada that is unaware of being stalked by a praying mantis, which is in turn the unconscious target of a strange magpie. As Zhuangzi lifts his bow to shoot the bird, he is startled by the warden of the forest who then pursues him. Although by no means a direct criticism of the Confucian conception of righteousness, this story of visual perspective highlights the precarious dependence on context of what is the right thing to do in any situation.

[6]And this is not to say that the knowledge is of a propositional kind (knowledge that). Indeed, ritual knowledge is typically primarily procedural (knowledge how), although it may also have a propositional component. For example, Confucius is reported to have bowed before ascending the stairs to the hall, perhaps automatically as a result of years of habituation, but he was also capable of reflecting on his actions and describing them.

affairs are not completed, ritual and music will not flourish. If ritual and music do not flourish, punishments and penalties will miss their mark. When punishments and penalties miss their mark, people lack the wherewithal to control hand and foot. Hence a gentleman's words must be acceptable to vocalize and his language must be acceptable as action. A gentleman's language lacks anything that misleads - period.

<div align="right">Analects 13.3[7]</div>

In addition to taking from this a recognition of the appreciation in ancient China of the immense power of names, we want to focus on the explicit connection between the application of names in speech and in action. Knowing a name, or better, being able to use a name correctly, is not just a matter of recognizing those things that it describes but responding to them with appropriate action. Names, *míng*, are *lǐ* in a microcosm: the humanly applicable fragments of the Confucian way. [8]

The way in which names guide actions is clearest in the case of social role names. Take 王 *wáng* (king) as an example. By recognizing his king as his king, the loyal subject will act in certain ways appropriate to his role as subject. The action may be anything from a simple gesture such as bowing to a complicated act of self-sacrifice. Take the simplest case. The king gives an audience to his minister, who bows when entering the king's presence. In this situation, the king is being kingly (or 南面而立 "facing south") and the minister is bowing. These two responses express the relationship between the two people: king to subject. [9]

Let us now turn to a simple model for this scenario, continuing with our earlier view of situations and responses. Suppose that in some situation s something x deserving of a particular name n is recognized as such by someone y. What is recognized by y is a certain action or response r_x to the situation by x which makes the classification of x by n acceptable. And y's recognition of this fact is also in the form of a response r_y. To capture a little of this in our model, we will associate with each *míng* (name) a set of triples of the form $\langle s, r_1, r_2 \rangle$, where r_2 is a response to r_1 conforming to a correct understanding of the name, as justified

---

[7]子路曰 :"衛君待子而為政，子將奚先？"子曰 :"必也正名乎 !"子路曰 :"有是哉，子之迂也！奚其正？"子曰 :"野哉由也！君子於其所不知，蓋闕如也。名不正，則言不順；言不順，則事不成；事不成，則禮樂不興；禮樂不興，則刑罰不中；刑罰不中，則民無所措手足。故君子名之必可言也，言之必可行也。君子於其言，無所苟而已矣。"《論語·子路》Translation by C. Hansen, *A Daoist Theory of Chinese Thought: A Philosophical Interpretation*, New York: Oxford University Press, 1992.

[8]Names are a very general term here, referring to any sort of linguistic category appropriate to the occasion: common nouns, verbs, and others.

[9]This point again illustrates our earlier decision to treat situations as consisting of responses. While not responding to the minister, perhaps, the king still occurs in the situation in a way that constitutes an appropriate response: acting in a king-like manner.

by the response r_1.[10] In other words, we do not make names a separate category that may, or may not be associated with certain appropriate behaviour: in a strong sense, the name, used correctly, is given by that behaviour.

For example, a situation s in which a minister confronted with his *wáng* (king) facing south r_1 responds by r_2 killing him, clearly represents an incorrect application of the name 王 *wáng*, in our terms: $\langle r_1, r_2 \rangle \notin wang_s$. It is important that this can be a two-way street: both r_1 and r_2 are crucial parameters of this model of names. Mencius tells us that he has never heard of a king being killed by his subjects but he has heard of a guy called Zhou (紂) being executed. [11] The implication is that if a man in a kingly position who does not act like a king, he is not a king and so killing him does not go against the ritual prohibition of regicide.

We can then model the idea presented in Analects 13.3 by taking the responsibility of government to be to ensure that everyone conforms to a set Z of correct uses of names.

> **Name structure** A *name structure* $\langle R, S, Z \rangle$ consists of a set R of *responses*, a set S of subsets of R (of *situations*) and a set Z of *names*, each of which is a set of triples $\langle s, r_1, r_2 \rangle$ where $r_1, r_2 \in s$ and $s \in S$.

The Utopian society envisaged in the above quotation is one in which everyone's responses to each other conform to Z. An obvious question then arises: what is the relationship between Y (modelling *yì*, right) and Z (modelling *zhèng míng*, correcting names)?

In the ideal case, we might imagine that they are equivalent in the following sense:

$$Y = \{\langle s, r \rangle \mid \langle s, r', r \rangle \in M \text{for some } r' \in R \text{ and } M \in Z\}$$

[12] In other words, an action r in a situation s is *yì* iff it is an appropriate response to some other action r' in s according to the correct use of some name M in Z. That is to say, the system of government is so perfect that the mere fact that the use of names has been put into order ensures that people always do what is right. This seems a tall order even for a Utopia. A little more realistically, we might require that

$$\{\langle s, r \rangle \mid \langle s, r', r \rangle \in M \text{ for some } r' \in R \text{ and } M \in Z\} \subseteq Y$$

---

[10]There is also an aspect of temporal succession in this senario which we will not model here.

[11]聞誅一夫紂，未聞弒君也。《孟子·梁惠王下》

[12]Formally this definition leaves room for conflicts where r might be a correct naming for one r' and at the same time for r'', howeover, the priority rules of classical chinese thinking makes it very unlikely.

This would say that using names correctly is always a right thing to do, a form of "soundness" of the social system. More ambitious would be a requirement of "completeness":

$$Y \subseteq \{\langle s, r \rangle \mid \langle s, r', r \rangle \in M \text{ for some } r' \in R \text{ and } M \in Z\}$$

where each correct response is triggered by a correct use of a name. In other words, then in any situation, if a response is *yì* then it conforms to *zhèng míng*. [13]

Let us pause to summarize the previous two sections. We started with a core Confucian notion, *yì*, and proposed an interpretation using "response structures". Appropriate responses under certain situations are considered as *yì*. We then extended our treatment of response structures to "name structures" in an attempt to explain (in the most general terms) the Confucian doctrine of *zhèng míng*, the idea that correcting names is an effective way of ensuring social harmony. Relationships between *yì*, *lǐ* and *zhèng míng* were discussed from a logical point of view. Admittedly, the models proposed here are very behaviourist, but then, so it classical Chinese philosophy.

From the next section, we will shift our attention to the Mohists, concentrating on their theory of names and argumentation. Although some connections between these topics will be apparent, we do no intend to make any direct bridge between the models considered so far and those of the rest of the paper.

4   From names to kinds

A concern with the relationship between *míng* (names) and *shí* (object, reality) is by no means limited to the Confucians. Discussion on this issue was found in almost every philosophical text of the period. In particular, it was one of the central themes of the Mohists. Instead of resorting to name rectification to regulate right behaviour, the Mohists took a different approach. They considered names together

---

[13]A possible counterexample to this strong version is given in Mencius 4A.17 " That men and women should not touch in handing something to one another is the ritual, but if your sister-in-law is drowning, to pull her out with your hand is discretion" (translation by Bryan van Norden in *Readings*, ibid). The pulling out is the right thing to do, but it misapplies the name of sister-in-law. That we are able to exercise discretion *qúan* (權) is therefore at least a potential problem for our somewhat simple model of Analects 13.3. This ties us, of course, with the general issue in deontic reasoning that norms can be defeasible under unusual circumstances. We will not pursue this line of thought here, although defeasibility might have interesting implications for a theory of correct naming.

with objects and they believed that their "mating" is key to define names. In the *Canons*, it says that

> 所以谓，名也。所谓，实也。名实耦，合也。
> What something is called by is its 'name'. What is so called is the 'object'. The mating of 'name' and 'object' is 'relating'. [14]

The Mohist canons 《墨經》 are an elaborate and systematic study of the significance of the names of particular philosophical importance, and even 名 *míng* itself is listed and is divided into three categories: unrestricted *dá* (達), classifying *lèi* (類), and private *sī* (私).[15] These three categories are explained by means of examples: thing *wù* (物), horse *mǎ* (物), and Zāng (臧), a personal name.

It it the second of these categories, 'classifying names' *lèimíng*, that we will be focus on for the remainder of the paper. We will now have to design models that look more deeply into the structure of naming, differently from the response-style definition that we considered before for the Confucian theory of names.

The full explanation of the use of names of the *lèi* 'classifying' type runs as follows:

> 命之「馬」，類也。「若實」也者，必以是名也。
> Naming something 'horse' is 'classifying' - for 'like the object' we necessarily use this name.[16]

We will interpret this definition very literally. To use the name 'horse' *mǎ* (馬) in a situation is to identify something as 'like the object' *rùo shí* (若實) . There is clearly a major new idea here which we need to bring out in our logical model, namely, that of *similarity*.This can be modelled using a binary similarity relation S on a domain D of objects. [17] Some of these objects will have been already classified as *mǎ* (horse), perhaps in the style of our earlier sections, say through an appropriate response. We write $a \models P$ to indicate that the object a has been classified as P. The explanation of *lèimíng* shows how classification can be extended to other objects. Each object a in the domain determines a *kind*, namely the kind of thing that is similar to a. For example, suppose D is the domain of animals and

---

[14]Translation by A.C. Graham in *Later Mohist Logic, Ethics and Science*, Hong Kong: The Chinese University, 2003, A81, p.328. Subsequent translations from the Mohist texts will all be by Graham. We will also follow his punctuations and textual emendations unless otherwise noted.

[15]Graham, A79, p.326.

[16]Graham, A78, p.325.

[17]For simplicity's sake, we will suppose here that it is 'objects' in some primitive sense that are being classified, although, as indicated in previous section, a more complete account would extend this to the classification of our earlier situation-action pairs.

S is the relation 'same shape'.[18] Then each horse h determines the kind of animal that has the same shape as h. Now for S to be a relation of 'similarity' we can assume that it is both reflexive (everything is similar to itself) and symmetric (if Qu Huang similar to Dapple then Dapple is similar to Qu Huang). [19] An obvious further question to ask is whether the similarity relation should also be taken to be transitive: Sab and Sbc implies Sac. If it is, then the division into kinds is a partition of the domain: every object is of a unique kind. This is a reasonable assumption for basic animal kinds, given a suitably flexible interpretation of 'shape.' [20] But it fails for other means of comparison, notably anything that admits of fuzzy boundaries.[21] If we compared objects of the basis of colour, for example, transitivity would lead us to the familiar contradictions of the Sorites Paradox.

There is further discussion of similarity as the basis of classification into kinds elsewhere in the Mozi:

止類以行之。說在同。
彼以此其然也說是其然也。我以此其不然也疑是其然也。

Fix the kind, in order to 'make the man proceed'. Explained by: the sameness.
The other, on the grounds that it is so of the instance here, argues that it is so of the thing it is; I, on the grounds that it is not so of the instance here, doubt that it is so of the thing it is. [22]

Here we see the earlier-discussed connection to action explicitly mentioned in the definition of 'fix' *zhǐ* (止). The classification of situations by kind determines how a person acts/proceeds *xíng* (行). And this is explained by similarity *tóng* (同). What follows the definition are examples of the use of 'similarity of kinds' *lèitóng* (類同) in debate. If objects (or situations) a and b are of the same kind then certain inferences are licensed. If they are both horses, for example, you might infer from

---

[18]Gonsun Longzi hints at this method of comparison in the famous 'White Horse' dialogue 《白馬論》 when he says that 'horse' is used to name the shape/form *xíng* (形); 'white' is used to name the colour.

[19]Qu Huang (渠黄) was a famous horse in the stables of King Mu of the Western Zhou (周穆王 956-918 BCE). Dapple was ridden by Sancho Panza in *Don Quixote*. Since Dapple is a mule neither animal is similar to the other, their obvious dissimilarity being the length of their ears. Dapple would therefore not be classified as *mǎ* (horse).

[20]Of course, we are concerned only to represent a pre-scientific classification of animals - the complexities of biological definitions of 'species' are quite beyond the scope of the present discussion.

[21]See K. Fine, Vagueness, Truth and Logic. *Synthese* 30 (3-4):265-300, 1975.

[22]Graham, B1 p.348.

a's having a tail shorter than its leg, that b's tail is also shorter than its legs. Conversely, I might infer from a's lack of canine teeth than b does not have canines. 'Sameness' *tóng* (同) and its antonym 'difference' *yì* (異) are also defined in the Canons in an obvious parallel:

> 同。重，體，合，類。
> 異。二，不體，不合，不類。

> *tóng* (same). Identical, as units, as together, of a kind.
> *yì* (different). Two, not units, not together, not of a kind. [23]

Further, in the explanation of these definitions, *lèitóng* (similarity of kinds) is 有以同 "being the same in some respect" and not being of the same kind is 不有同 "not having what is the same." In this we see a recognition of the fact that things of the same kind are not identical in every respect, only in some. The inferences of the kind just mentioned must therefore restricted accordingly. We cannot, for example, infer from one horse being white that the other horse is also white. Whiteness is just not part of the *lèitóng* for horses. In the next section, we will look further into how the Mohists saw the underlying structure of similarity.

5 Standards and criteria

The Mohists developed the study of *lèitóng* further using the concepts of 'standards' *fǎ* (法) and *yīn* (因):

> 法，所若而然也。
> 意規員三也，俱可以為法。
> The *fǎ* (standard) is that in being like which something is so.
> The idea, the compasses, a circle, all three may serve as standard.

> 因，所然也。
> 「然」也者，貌若法也。
> The *yīn* (criterion) is that wherein it is so.
> Being 'so' is the characteristics being like the standard. [24]

In other words, to classify something as a *yúan* (circle), one must have some *fǎ* (standard) against which to compare it. This may be the idea of a circle, an instrument for drawing circles, or another circle. The third of these standards is

---

[23] Graham, A86, A87, p.334.
[24] Graham, A70, A71, p.316.

already represented in our model of objects with a similarity relation. We use an actual horse to judge that some other animal is also a horse, by comparison with the horse we have. It would be interesting to see how to add the other two types of *fǎ*, which raise interesting issues of their own, but we leave this matter for future investigation. Even so, we do need an extension of our model so far to deal with second new aspect, the notion of *yīn* (criterion), though so far scholars have not paid much attention to it. We make our attempt here. Earlier we made the simplifying assumption that there was just one relation of similarity. But in the texts, we see the Mohists articulating the need for methods of comparison that vary according to the kind. The 'Lesser Pick' *Xǐaoqǔ* 《小取》 chapter of the Mozi has some nice examples of why this is necessary:

> 之馬之目眇則謂之馬眇，之馬之目大而不謂之馬大。
> 之牛之毛黃則謂之牛黃，之牛之毛眾而不謂之牛眾。
>
> If this horse's eyes are blind, we say this horse is blind; though this horse's eyes are big, we do not say that it is a big horse.
> If these oxen's hairs are yellow, we say these oxen are yellow; though these oxen's hairs are many, we do not say that these oxen are many.
> [25]

Clearly, some aspects are thought to be crucial to defining kinds (say, eyes for a blind horse), but other aspects are not (say, eyes for a big horse). To model this we propose a simple account of judgements about kinds Y. In order to make claims of the form '$X\ Y$ 也' and 'X 非 Y 也,' there must be an agreed *fǎ* (standard) and *yīn* (criterion) for Y. We take these to be a name $fǎ_Y$ and a similarity relation $yīn_Y$, respectively. $fǎ_Y$ is a name for what might be termed a 'prototypical' Y and $yīn_Y$ is the *way or aspect* in which objects should be compared to a $fǎ_Y$. Both are necessary conditions for naming a kind.

> **Standard/criterion structure**   A *standard/criterion structure* consists of a domain D of situations/objects and a set of names, with a standard $fǎ_X$ and similarity relation $yīn_X$ on D, for each kind X. Now for an object to be of kind X it must stand in the relation $yīn_X$ to an instance of $fǎ_X$.

The judgement 'XY 也' is acceptable when there is some standard a for X and standard b for Y such that everything appropriately similar to a (according to the criterion for X) is also similar to b (according to the criterion for Y). In other words:

---

[25]Graham, NO18, p.492.

There are $a \models f\check{a}_X$ and $b \models f\check{a}_Y$ such that for every c in the domain, if $y\bar{\imath}n_X(c, a)$ then $y\bar{\imath}n_Y(c, b)$.

As a simple example, the judgement 'white horse is horse' 白馬馬也 [26] is acceptable because there is a horse s that we have classified as being white, which serves as a standard for being a white horse, so $s \models f\check{a}_{白}$, and a horse h that is a standard for being a horse $h \models f\check{a}_{馬}$ and every other animal a that is similar to s according to the criterion for white horse (which we may suppose is having the same colour and same shape), so $y\bar{\imath}n_{白}as$, is also similar to h according to the criterion for horse (merely having the same shape), that is $y\bar{\imath}n_{馬}ah$.

We think that this stipulation stays close to the material in our texts, even though it also has some surprising features. [27] As for formal properties, the relation between predicates defined in this way is reflexive, but not symmetric, and not transitive (since we did not assume transitivity for the underlying similarity relations). One might expect that the relation is 'monotonic' in allowing for predicate extension for X, Y, but again, this will depend on how the similarity relations for predicates and their extensions are connected.

In this section we have shown how a simple model can be given for *lèi*-based analogy reasoning in terms of binary relational structures with predicate-dependent prototypes and similarity relations. We do not claim that our concrete proposal fits every aspect of Mohist reasoning based on kinds: in particular, our stipulation for 'X, Y 也' judgements admits of alternatives that would quantify differently over available prototypes for X and Y. But whichever way we do this precisely, our point is that there is no deep dividing line with developments inside formal logic here, but rather strong analogies.[28]

6   Argumentation in the Mohist tradition

Our final topic concerns another major theme of the Mohists, perhaps even their primary concern: the process of argumentation itself, bringing out the different actors behind the scenarios that we discussed earlier. Here we recall the definition

---

[26] 白 is a name for 'white' , 馬 is a name for 'horse' and 白馬 is a name for 'white horse'.

[27] For instance, showing that 'X, Y 也' does *not* hold would involve looking at all prototypes for both X and Y.

[28] Similarity relations of many kinds (transitive, or not) have been studied in many formal disciplines, from Measurement Theory (E. Pap, ed, *Handbook of Measurement Theory*, Amsterdam: Elsevier, 2011.) to the semantics of conditionals (cf. D. Lewis, *Counterfactuals*, Oxford: Blackwell, 1973) or vague predicates (R. van Rooij, Vagueness and Linguistics. In G. Ronzitti ed, *The Vagueness Handbook*, to appear.).

of disputation:

> 辯，爭彼也。辯勝，當也。
> *bìan* (disputation) is contending over claims which are the converse of each other. Winning in disputation is fitting the facts.[29]

> 謂辯無勝必不當，說在辯。
> To say that there is no winner in *bìan* necessarily does not fit the facts.[30]

In particular, the first quote implies that one doesn't (or shouldn't) lose the dispute merely because one has a controversial opinion. The 'facts,' which is Graham's translation of *dàng* provide an anchor for reason's otherwise slippery progress. The way in which this appeal to objectivity provides a norm for disputes is given in the explanation of these two canons:

> 辯：或謂之「牛」，或謂之「非牛」，是爭彼也。是不俱當，不俱當必或不當。(不若當「犬」)。
> One calling it an 'ox' and the other 'non-ox' is 'contending over claims which are the converse of each other'. Such being the case they do not both fit the fact; and if they do not both fit, necessarily one of them does not fit. Not like fitting 'dog'.[31]

> 謂：所謂，非同也則異也。同則或謂之狗，其或謂之犬也，異則或謂之牛，其或謂之馬也，俱無勝，是不辯也。「辯」也者，或謂之是，或謂之非，當者勝也。
> The things that something is called are different. In the case where they are the same, one man calling it 'whelp' and the other man a 'dog', or where they are different one calling it an 'ox' and the other a 'horse', and neither winning, is failure to engage in *bìan*. In '*bìan*', one says it is this and the other that it is not, and the one who fits the facts is the winner. [32]

The gist of this is that for a dispute to count as a proper *bìan*, the disagreement must come down to a matter of fact, with the primary example being the categorisation of objects. To argue about whether something is 牛 (an ox) is legitimate because one side says it is, the other side says it isn't and only one of them can be right.

---

[29] 《墨子·經上》Graham, A74, p.318.
[30] 《墨子·經下》Graham, B35, p.402.
[31] 《墨子·經說上》Graham, A74, p.318.
[32] 《墨子·經說下》Graham, B35, p.403.

If the dispute is about whether something is 狗 or 犬 (two words for 'dog') is it not a proper *bìan*, since no matter of fact is available to make one side the winner and the other the loser. The canonical case of *bìan*, for the Mohists, is a dispute in which there is a clear criterion of success, and much of the Canons can be seen as an attempt to provide such criteria. We will focus on the simplest of these: the *lèi* categorisation judgements discussed earlier. For these, as we have seen, there is an attempt to provide a theory of what is involved, one that we have already modelled using "standard-criterion structures". So suppose I claim that x is 牛 (an ox), and you disagree. To do so, you have to provide a standard and criterion for ox according to which x is 非牛 (not an ox). Now, in the case of 牛, the standard and criterion are presumably common ground but we will continue as if they were not, so as to see how the debate can continue. In out later discussions we will also the simplifying assumption that the standard for 牛 is an actual ox, y. How to model this argumentation scenario? There is a good deal of literature on the logical aspects of these and other Mohist examples, pointing at various analogies between Mohist views of disputation and the basic laws of propositional logic, viewed as discourse principles.[33] We have nothing to add to this literature, but instead, we will put our modeling approach to work on another, perhaps more ambitious project.

An argumentative practice usually exists, but it is not written down completely, the way we have a score for a dance performance, but perhaps not a complete record of all the crucial moves. This does not mean that this practice was not as important as the preserved texts. Compare the case of medieval Western logic where there is a well-attested tradition of argumentation games [34] making the logical tradition much richer than just the extant texts on the syllogistic and related topics. What we know is that these games were often about consistency management, challenges were issued, and players would have to commit to some stated assertion, whether in the affirmative of the negative. Now it is hard to model a practice, for which we have only very incomplete traces. [35] Still, it seems of interest to make some attempts, and indeed, formal modeling is still useful. In particular, there is a long tradition of "logic games" modeling various kinds of debate. [36]

---

[33]Cf. D. Leslie. Argument by contradiction in pre-Buddhist Chinese reasoning, Australian National University, Canberra, 1964.

[34]Cf. C.L.Hamblin, *Introduction to Medieval Logic*, Oxford University Press, 1970.

[35]We do think that it would be very important to bring together all that is known on ancient Chinese intellectual practices such as philosophical debate, but also legal disputation, rhetoric, and perhaps other sources.

[36]Cf. J. van Benthem, *Logic in Games*, manuscript, 2011 and W. Hodges, *Building Models by Games*, Cambridge University Press, 1985.

In the remainder of this paper, we develop one such game, staying close to the similarity structure discussed in the previous sections, just to give an impression of what it means to formally model a practice. [37]
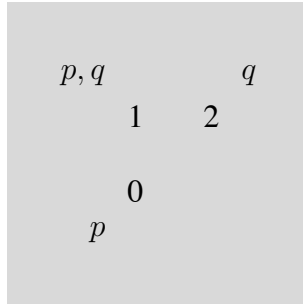
7   Argumentation games

In this section we will explore how our models can be used to analyze an argumentative practice. Our proposal is just a "proof of concept" for a direction that we think can be profitably applied to studying Mohist logic. In what follows, we present the main features of an argumentation game and illustrate our ideas by examples.

First, the facts are modeled by a structure in the spirit of our Section 4, that we will call a *classification*, defined as

> A **classification** is a tuple $F = \langle W, M, \models \rangle$ where W is a set of things, M is a set of names, and $w \models m$ represents the judgement that w is correctly described as m.

A classification can be pictured as a diagram in which we use numbers to stand for the things in W and letters to stand for the names in M. For example,

$$p, q \qquad\qquad q$$
$$1 \qquad 2$$
$$0$$
$$p$$

---

[37]Another candidate for this sort of modeling would be *argumentation theory*, which has been developing quite rigorous connections with logic recently (D. M. Gabbay, R. H. Johnson, H. J. Ohlbach, and J. Woods, eds, *Handbook of the Logic of Argument and Inference*, Amsterdam: North-Holland, 2002 and P.M. Dung, On the Acceptability of Arguments and its Fundamental Role in Non-monotonic Reasoning, Logic Programming and N-person Games, *Artificial Intelligence* 77:321–357, 1995.) Yet one more relevant research line are current dynamic logic of information flow in interactive scenarios, cf. J. van Benthem, *Logical Dynamics of Information and Interaction*, Cambridge University Press, 2011 and F. Liu, *Reasoning about Preference Dynamics*, Springer, 2011.

represents the classification of things $W = \{0, 1, 2\}$ with names $M = \{p, q\}$ so that $0 \models p$, $1 \models p$, $1 \models q$ and $2 \models q$ (and that's all).

We will assume that some classification F is common ground between the disputants. They are arguing about the extension of this classification to a new *lèi* (category) L. We want to account for various ways in which this can be done. The simplest is just to state that some thing w either is or is not L. Next is an inference from one thing u being L to another thing v being L on the basis of their similarity. But similarity can be challenged, and typically this is done by pointing out that u has some property that v lacks, or vice versa. Any such distinguishing property must be expressible by a name in the set M in order for it to be used in the dispute. We need to keep a careful check on the disputants' commitments in each of these three areas: the extension E of L, the relevant similarity relation $\sim$, and those property (names) P that can distinguish things that are L from things that are not.

We will now define a few basic notions reflecting these features and after that we will be in a position to define our game.

Categorisation structures   At any state of the dispute, we model each disputant's claims about L as a **categorisation structure** $K = \langle F, E, \sim, P \rangle$[38] in which F is the agreed classification and

1. E consists of two sets E^+ and E^-. E^+ represents the set of things that the disputant has claimed to be L and E^- represents the set of things that the disputant has claimed to be not L (*not L*).

2. $\sim$ consists of two binary relations $\sim^+$ and $\sim^-$, where $u \sim^+ v$ represents the disputant's having claimed that u is similar to v and $u \sim^- v$ represents the disputant's having claimed that u is *dis*similar to v. Similarity here should be understood as the notion of similarity that is relevant to L-categorisation.

3. P is a set of names (in M) representing those features that the disputant has claimed to be relevant to L-categorisation. A complete set P of such names would provide a criterion for similarity in L-categorisation, so that sameness with respect to each m in P would a necessary condition for similarity, and sameness with respect to every m in P would be sufficient.

Categorisation structures, too, can be pictured. We represent similarity judgements by lines connecting the things that are deemed similar or dissimilar. Similarity $\sim^+$

---

[38]This follows a long tradition in games and discourse studies of listing players' current commitments. cf. e.g., D. N. Walton and E. C. W. Krabbe, *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, Albany: State University of New York Press, 1995.

is represented by a solid line and dissimilarity $\sim^-$ by a dashed line. The elements of P will be listed at the top of the diagram. [39] For example,
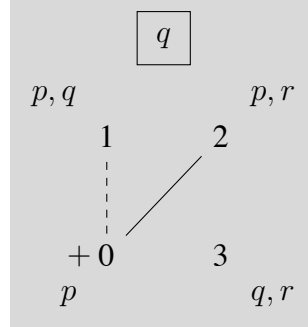
In this categorisation structure, our disputant has committed himself to saying that 0 is L and that 0 is similar to 2 but dissimilar to 1, and also that q names a feature that is necessary for similarity (with respect to L). In other words, any two things that are said to be similar (now or in the future) must have the same q-value: they must either both be q or both not q. This severely constrains his future options. Thing 2, for example, must eventually be conceded as being L and thing 3 cannot be taken as similar to 0 or to 2.

Maintaining consistency   The set P defines an equivalence relation on W:

$$u \approx_P v \quad \text{iff} \quad \text{for all } p \in P, u \models p \text{ iff } v \models p$$

This is the similarity relation implied by taking P to define similarly. It need not be the same as $\sim$, but we require consistency in the following sense. Say that a categorization strucure K is *consistent* iff for all $u, v \in W$:

1. If $u \sim^+ v$ then $u \approx_P v$, i.e., if two things are claimed to be similar (with respect to L) then they must have the same P-features. This captures the individual necessity of the features named in P for L-similarity.

2. If $u \sim^- v$ then $u \not\approx_P v$ i.e., if two things are claimed to be dissimilar with respect to L then there must be a P-feature on which they differ (so that one has the feature and the other lacks it). This captures the joint sufficiency of the features named in P for similarity with respect to L.

3. If $u \in E^+$ and $v \in E^-$ then $u \not\sim^+ v$, i.e., if one thing is claimed to be L and another to be not-L then they cannot be claimed to be similar (with respect to L).

---

[39] Note also that these structures are standard three-valued Kripke models.

4. If $u, v \in E^+$ or $u, v \in E^-$ then $u \not\sim^- v$, i.e., if two things are both claimed to be L or are both claimed to be not-L, then they cannot also be claimed to be dissimilar with respect to L.

These are the natural requirements that one would expect for a consistent usage of the notions of similarity and predication. Therefore, people can be held to them, and they can be used as the basis for an account of disputation. [40]

With all this in place, we can now define our game.

Game state   For simplicity, we will suppose that there are two disputants, and so the state of the game at a given stage is represented by a pair of consistent categorisation structures $\langle K_1, K_2 \rangle$. The first categorisation structure K_1 represents the commitments of the first player and K_2 represents those of the second. The *initial* game state for a given common classification F is just $\langle K_\emptyset, K_\emptyset \rangle$ where $K_\emptyset = \langle F, E_\emptyset, \sim_\emptyset, P_\emptyset \rangle$ with $E_\emptyset, \sim_\emptyset$ and $P_\emptyset$ all empty.[41] This represents a starting point in which the disputants shared classification is F and in which neither has made any claim about the new category L. In case there is some initial agreement between disputants about L, represented by K, we will see that a single move of the game can take us to state $\langle K, K \rangle$.

An utterance made by one of the disputants has the effect of extending that disputant's categorisation structure so that it represents the new claim. The utterance may be about any aspect of the categorisation: that a particular thing is or is not an ox, that one thing is or is not similar to another (with respect to ox-ness), or that being similar requires some specific feature, such as the same kind of teeth. Any new claim that is made must of course be consistent with what the disputant has previously claimed, and so there is a disincentive to say too much. The benefit of making a new claim is that the other disputants are required to respond, either agreeing or disagreeing with what is said. It is this tension between maintaining consistency and forcing others to respond that gives life to the game.

Two argumentation relations   To model the effect of making a new claim, we define two relations between categorisation structures. The first is that one structure $K' = \langle F', E', \sim', P' \rangle$ is an *extension* of a previous structure $K = \langle F, E, \sim, P \rangle$, written $K' \geq K$, which holds just in case $F' = F$, $E'^+ \supseteq E^+$, $E'^- \supseteq E^-$, $\sim'^+ \supseteq \sim^+$, $\sim'^- \supseteq \sim^-$, and $P' \supseteq P$. This is just to say that every claim represented by K is preserved in K', with the possible addition of new claims.

---

[40] This analysis of consistency in *lèi*-based predication may be seen as providing more substance to our formal account proposed in Section 5.

[41] More precisely, $E_\emptyset^+, E_\emptyset^-, \sim_\emptyset^+, \sim_\emptyset^-, P_\emptyset$ are all $\emptyset$.

The second relation requires some more definitions. We say that the *domain* of the categorisation relation E is $\mathrm{Dom}(E) = E^+ \cup E^-$. This represents the set of those things that have been categorised in K, either as L or as not L. Similarly, the domain of $\sim$ is $\mathrm{Dom}(\sim) = \sim^+ \cup \sim^-$. The domains of these parts of a categorisation structure determine how complete the categorisation is. In the limiting case, starting with a classification $F = \langle W, M, \models \rangle$, the categorisation structure $K = \langle F, E, \sim, P \rangle$ is *complete* when everything is categorised, $\mathrm{Dom}(E) = W$ and every pair of things are judged similar or dissimilar, $\mathrm{Dom}(\sim) = W \times W$.[42] Now, in a debate we will require that each disputant can respond to the other's claim, in the sense that his categorisation structure is at least as complete as his opponent's. To express this, we say that $K_1 = \langle F_1, E_1, \sim_1, P_1 \rangle$ is *at least as complete as* $K_2 = \langle F_2, E_2, \sim_2, P_2 \rangle$, written $K_1 \gtrsim K_2$, just in case, $F_1 = F_2$, $\mathrm{Dom}(E_1) \supseteq \mathrm{Dom}(E_2), \mathrm{Dom}(\sim_1) \supseteq \mathrm{Dom}(\sim_2)$.

Moves of the game   Finally, we are in a position to say what it is for a disputant to make a legitimate move in the game. A *move* consists of a change of game state $\langle K_1, K_2 \rangle$ to a $\langle K_1', K_2' \rangle$ such that either

1. $K_1' \geq K_1$ and $K_1' \gtrsim K_2$ and $K_2' = K_2$ (a move by the first player), or

2. $K_2' \geq K_2$ and $K_2' \gtrsim K_2$ and $K_1' = K_1$ (a move by the second player).

In other words, play proceeds by one of the players extending his categorisation structure in any way that is itself consistent, and which is at least as complete as his opponent's categorisation structure.[43]

A few special cases are worth mentioning. If the disputants are in complete agreement, so that $K_1 = K_2$, then the game state $\langle K_1, K_2 \rangle$ is stable in the sense that the trivial move of changing nothing is a legitimate move of the game. A similar situation also arises even if $K_1 \neq K_2$ in the case that $K_1 \gtrsim K_2$ and $K_2 \gtrsim K_1$. For then, the claims of the two disputants, although they do not agree, are equally complete, and so neither of their claims requires a response from the other. [44]

Note that although we do not insist on the disputants alternating their moves, in practice, we may as well have, since any sequence of two moves by the same

---

[42] Note that the consistency conditions ensure that for a complete and consistent categorisation structure, P provides a complete definition of similarity in the sense that that $\approx_P$ and $\sim^+$ coincide.

[43] The condition that $K_2' = K_2$ in the case of a move by the first player merely enforces the obvious condition that the first player can only directly effect her own categorisation structure.

[44] As previously mentioned, initial agreement between disputants, represented by a consistent categorisation structure K, can be modelled as sequence of moves in which each player declares the common structure: the first player moves from $\langle K_\emptyset, K_\emptyset \rangle$ to $\langle K, K_\emptyset \rangle$ and then the second moves to $\langle K, K \rangle$ (or vice versa).

player can be achieved in a single move. For example, if the first player moves from $\langle K_1, K_2 \rangle$ to $\langle K_1', K_2 \rangle$ and then to $\langle K_1'', K_2 \rangle$, she may as well have moved straight from $\langle K_1, K_2 \rangle$ to $\langle K_1'', K_2 \rangle$. There will be no effect on the game.

Winning and losing   Finally, we define what is to win or to lose an argumentation game. A game state is a *loss* for a disputant (and a *win* for his opponent) if there is no move that he can make. A loss for a disputant is a situation in which there is no way for him to respond to his opponent in a way that is consistent with what he has said before. This is our model of 無當 ('failing to fit the facts'). A *game* is a sequence of game states, $\langle K_1^0, K_2^0 \rangle$, $\langle K_1^1, K_2^1 \rangle$, $\langle K_1^3, K_2^3 \rangle$, etc. that starts with the initial game state, $K_1^0 = K_2^0 = K_\emptyset$. It may finish after a finite number of moves, n, when $\langle K_1^n, K_2^n \rangle$ is a loss for one of the disputants (and so a win for the other), or continue indefinitely. In the latter case, if the common classification F is finite (meaning that the set W of things to be classified is finite), the argument must eventually reach a certain stage n such that K_1^n and K_2^n are both complete and the remaining moves of the game are the trivial moves of doing nothing.[45] In this case, the game is a *draw*.

A concrete illustration   These concepts can be illustrated by the following simple example, which although not taken directly from an ancient Chinese source is loosely based on the example of *biàn* given in the Mozi and quoted earlier: "One calling it an 'ox' and the other 'non-ox' is 'contending over claims which are the converse of each other'. Such being the case they do not both fit the fact; and if they do not both fit, necessarily one of them does not fit."[46]

A: It's an ox.
B: No, it isn't.
A: But that's an ox. (Implicit: they are similar).
B: Yes, it is, but it has horns. (Horn-possession is a distinguishing feature.)
A: Yesterday, we saw an ox without horns. (Implicit: but similar to this one, in being an ox)
B: (Loses)

Let's suppose that the common classification consists only of things mentioned in the dispute. Call the thing they are arguing about 0. Call the ox mentioned by A for comparison 1. And call the ox they saw yesterday, 2. The only feature

---

[45]There is also the possibility of extending P, but since the two categorisation structures are both complete, this has no effect on the outcome of the game. Any move by one player to extend P can be ignored by the other (who can make a trivial move). If the set W of things to be classified is infinite, some other possibilities emerge; but we will not consider these here.

[46]或謂之「牛」，或謂之「非牛」，是爭仮也。是不俱當，不俱當必或不當。《墨子・經說上》Graham, A74, p.318.

mentioned in 'having horns', which we will represent as h, but in anticipation of a later example, we will also include the feature c of 'having canines'. We will suppose that the common classification F consists of a set $W = \{0, 1, 2\}$ of things, a set $M = \{h, c\}$ of 'names' and the relation $\models$ that holds between 1 and h, 0 and c (i.e., $1 \models h$ and $0 \models c$) and nothing else. This can be pictured as follows:
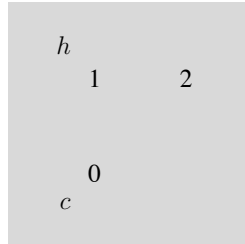
In the course of the dispute, what is at issue is how to assign the property "ox" to the objects in this struture, as well as which further predicates are relevant to this determination.

For a start, there is an appeal to 'the ox we saw yesterday,' so this, at least, must also be commonly agreed. So let K be the categorisation structure $\langle F, E, \sim, P \rangle$ which is just like $K_\emptyset$ but with 2 categorized as an ox, so that $E^+ = \{2\}$. We can represent the fact that K is common to both disputants by supposing that the first two moves are to $\langle K_\emptyset, K \rangle$ and then $\langle K, K \rangle$. The dispute then proceeds through a sequence of moves. We will display only B's categorisation structure after the completion of each of his moves:
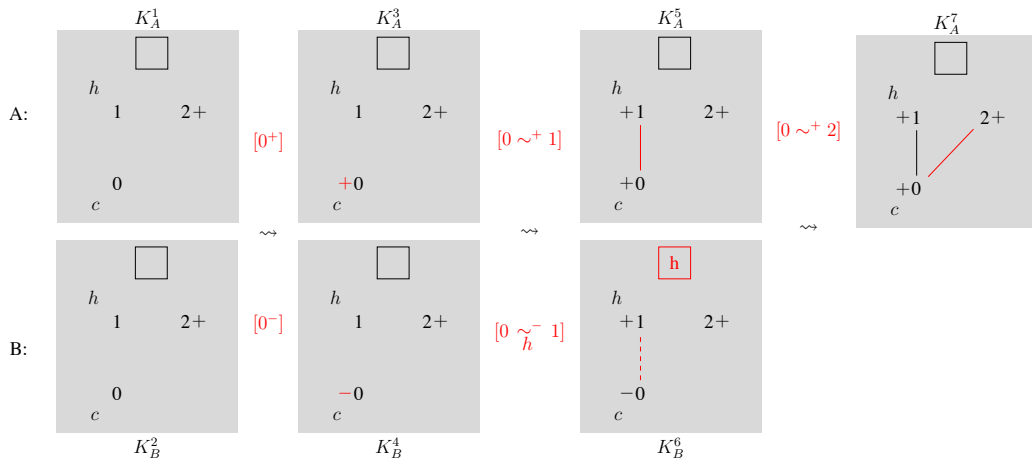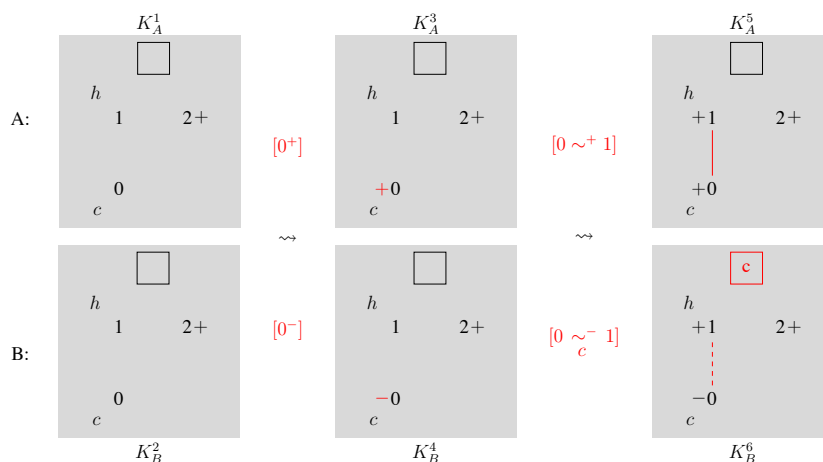

The pair of diagrams to the left depicts the commonly agreed categorisation $K_A^1 = K_B^2 = K$. A says "it's an ox," referring to 0, giving the categorisation K_A^3, which only differs from K_A^1 in the addition of 0 to E^+. B responds with "no, it isn't,"

giving the categorisation K_B^4. Although A's move is voluntary, in the sense that he could have stayed with K_A^1, B is forced to respond because $K_B^2 \not\gtrsim K_A^3$. For instance, the domain of E in K_A^3 contains 0, which is not contained in the domain of E in K. In other words, B is compelled to give his opinion on 0. B's move does not force A to make a further move. The game could end here, with neither player making a further claim. This would be a draw.

But A continues with "but that's an ox," referring to 1, and we take him to be making a further implicit claim that 0 is similar to 1. This extends his categorisation structure to K_A^5. The domain of both E and $\sim$ have been extended, and so B must respond. He responds with "yes, it is, but it has horns," agreeing to the addition of 1 to E^+ but disagreeing about the addition of $\langle 0, 1 \rangle$ to $\sim^+$. This gives him K_B^6, shown above with the dashed line indicating that 0 is dissimilar to 1. Consistency requires him to make a further extension, adding the name h, which we take to express horn-possession, to the set P. Finally, A reminds B of the ox 2 that they saw yesterday, making an implicit comparison with 0, ending with a categorisation structure K_A^7, to which B cannot respond. There is no structure K with $K \geq K_B^6$ and $K \gtrsim K_A^7$ because to ensure that $K \gtrsim K_A^7$, B would have to judge either that $0 \sim 2$ or that $0 \not\sim 2$. The first of these choices is inconsistent with $0 \in E^-$ and $2 \in E^+$; the second is inconsistent with $0, 2 \models h$ and $h \in P$. B's position fails to fit the facts in the game, and so he loses. Consequently, A wins.

Another scenario This is unfortunate for B but avoidable. If he had chosen a different distinguishing feature, B could have continued. Suppose, for example, the dialogue had proceeded in the following way:

A: It's an ox.
B: No, it isn't.
A: But that's an ox. (Implicit: they are similar).
B: Yes, it is, but it doesn't have canines. (Canine-teeth-possession is a distinguishing feature.)
A: ...

K_A^1

A:
h 1 2+
0
c

$[0^+]$

K_A^3

h 1 2+
$+0$
c

$[0 \sim^+ 1]$

K_A^5

h +1 2+
$+0$
c

$\leadsto$

K_B^2

B:
h 1 2+
0
c

$[0^-]$

h 1 2+
-0
c

$[0 \underset{c}{\sim^-} 1]$

h +1 2+
-0
c

c

$\leadsto$

K_B^4 K_B^6

This is a loss for neither disputant. A could continue with K_A^7, as before, but it would be met easily with K_B^8, in which B denies that 0 is similar to 2:

K_A^7

A: h +1 2+
$+0$
c

B: c
h +1 2+
-0
c

K_B^8
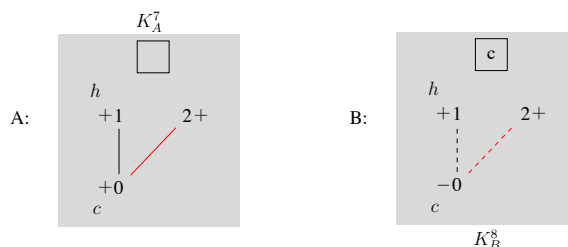
Discussion The particular Similarity Game defined here offers a style of thinking about discourse and argumentative commitment that borrows from similar traditions in the study of western logic.[47] What is important in this approach is the modeling of 'discourse flow' and 'commitment pressure' that are so typical of real debate. Of course, we are not committed to any particular 'nuts and bolts' of the game – and moves and winning conventions could in fact be modulated to capture various argumentation scenarios in the Mohist (or other ancient Chinese) texts. In particular, one feature that we have left open in our formulation of the game are possible restrictions on the sorts of objects and predicates that may legitimately be put on the table. Clearly, any argumentative practice needs constraints on what can be brought up, if one is to finish at all. [48]

There is much more to our game than we can present here. In particular, games typically allow for *strategic behaviour*: players can choose and plan their responses

---

[47]Cf. the classic *Dialogische Logik* by Lorenzen and Lorenz, 1978.

[48]There is also much to be said motivating the particular design of these games using Mohist ideas about inference, such as *tui*.

to what others do in advance, and they can analyse their total 'position' in the game. Indeed, many existing logic games have the following interesting property: the game is 'determined', meaning that one of the two players has a *winning strategy* that guarantees success against any possible play of the opponent.[49] Likewise, when we put a finite horizon in our similarity games, making it impossible to play for longer than some finite sequence of moves, Zermelo's Theorem will apply, showing that one of the players has the principled advantage of being able to force either a win or a draw. We think that this game feature is very close to the Mohist view of the aim of disputation, and it adds a dynamic flavour to the usual logic analysis of *bian*.

8   Conclusion

In this paper, we have taken a look at some key aspects of ancient Chinese views of language and argumentation, particularly those of the Confucians and Mohists. We have advocated the use of light-weight logical models of some key concepts, without importing the full machinery of modern logical calculi. In particular, we have suggested some models of the notion of discrimination and response, the idea of correct naming in terms of appropriate responses, and the extension of kinds via standards and criteria. We have also made a first attempt at representing the practice of argumentation using game models. Here the focus has been on games in which the categorisation of objects is extended in a way commensurate with Mohist ideas about the classification into kinds. The relationship between games and argumentation is an old theme in the study of logic in the West, and we believe that much can be learned by applying this methodology to the practices of ancient China. [50] Verification of this approach will require a closer study of textual material, of a kind that goes beyond the limited aims of the present paper.

Finally, we repeat a point that has been made at several places in this paper. As things stand now, the formal models that we have presented are just proposals illustrating a kind of approach. Clearly, judging their adequacy requires serious textual studies, of a kind that we have not provided here, but hope to start in follow-

---

[49]This applies in the Similarity Game when the categorisation structure of one player has a consistent and complete extension, but that of the opponent does not. In this case, the winning strategy is simple: if he continues extending until his structure is complete, he will win. If both players have categorisation structure that can be consistently completed, then either can force a draw. If, however, the categorisation structure of neither player can be consistently completed, some real gamesmanship is possible. We have yet to analyse the strategic aspects of such games.

[50]A similar approach to the disputes of mediaeval logicians has been taken recently by Catarina Dutilh Novaes (http://rug.academia.edu/CatarinaDutilhNovaes) and Sara L. Uckelman (http://staff.science.uva.nl/ suckelma/).

up work. But even at this stage, we believe that we have offered some tools and perspectives that may be worth adding to the ways in which the current community of scholars is trying to understand the fascinating phenomenon of Chinese logic two thousand years ago.