

# On the formalization Socratic dialogue

Martin Caminada  
Utrecht University

**Abstract:** In many types of natural dialogue it is possible that one of the participants is more or less forced by the other participant towards making commitments. This force usually comes in the form of sharp questions that guide a participant into a specific direction (like a contradiction). Examples of such dialogues occur in the philosophy of Socrates, critical interviews and the cross-examination of witnesses in court. In the current paper, we examine the characteristics of these kinds of dialogues, and provide a first sketch of how they could be formalized.

## 1. Introduction

Formal dialogue systems can provide a means for analysing informal dialogue (MacKenzie 1979, Hamblin 1970), as well as a basis for the design of agent communication protocols (Parsons et al 2003), or for the design of intelligent systems that can interact with humans (Milward & Beveridge 2004).

Dialogue systems like MacKenzie's DC (MacKenzie 1979) and Hamblin's H (Hamblin 1970) often focus on letting a player put forward a claim and then (possibly iteratively) letting him provide reasons for this claim. In natural dialogues, however, one also observes a different phenomenon. The idea is to let a player make a claim and then (possibly iteratively) to confront him with the consequences of this claim, forcing him to commit to statements that he may have wanted to avoid (such as contradictions). examples of such a discourse can be found in the philosophy of Socrates, the cross-examination of witnesses in court (Fulda 2000), critical interviewing techniques, as well as in modern philosophical argument (Skidmore 2002). We therefore regard the question of how such discourses can be formally modelled to be a relevant one.

## 2. Preliminaries

In the remainder of this paper we will follow (Pollock 1992,1995) in distinguishing two kinds of rules: *strict* and *defeasible*. A strict rule is an expression of the form  $\phi_1, \dots, \phi_n \rightarrow \Psi$  where  $\phi_1, \dots, \phi_n, \Psi$  are statements. Such a rule should be read as “If  $\phi_1$  and ... and  $\phi_n$  are the case, then it is *without any possible exception* the case that  $\Psi$ .” A defeasible rule is an expression of the form  $\phi_1, \dots, \phi_n \Rightarrow \Psi$  where  $\phi_1, \dots, \phi_n, \Psi$  are statements. Such a rule should be read as “If  $\phi_1$  and ... and

$\Phi_n$  are the case, then it is *normally* the case that  $\Psi$ .”

Using strict and defeasible rules, one can then start to construct (formal) *arguments*. An example would be the following.

### **Example: Red Object**

“*The object is red because John says it looks red.*”

Says(John, Looks(o, red))  $\Rightarrow$  Looks(o, red)

Looks(o, red)  $\Rightarrow$  Is(o, red)

An argument, like the one above, can be attacked basically in two different ways (Pollock 1992, 1995): by *rebutting* and *undercutting*. A rebutting attacker argues for the negation of one of the attacked argument's conclusions. An example would be the following.

### **Example: Blue Object**

“*The object is not red because Mary says it looks blue.*”

Says(Mary, Looks(o, blue))  $\Rightarrow$  Looks(o, blue)

Looks(o, blue)  $\Rightarrow$  Is(o, blue)

Is(o, blue)  $\rightarrow \neg$ Is(o, red)

An undercutting attack, on the other hand, argues merely against the applicability of a rule used in the attacked argument. An example would be the following.

### **Example: Red Light**

“*Suzy says the object is merely illuminated by a red light.*”

Says(Suzy, Illuminated(o, red\_light))  $\Rightarrow$  Illuminated(o, red\_light)

Illuminated(o, red\_light)  $\Rightarrow \neg$ [Looks(o, red)  $\Rightarrow$  Is(o, red)]

Notice that the fact that the object is illuminated by a red light is not a reason for inferring that it is *not* red. It merely indicates that the fact that it looks red is no longer a reason for inferring it actually is red. Thus, the relevant rule (Looks(o, red)  $\Rightarrow$  Is(o, red)) is undercut.

Whereas in formal argumentation, like (Prakken and Sartor 1997, Dung 1995), an entire argument is provided at each move, in formal dialogue (Prakken 2005, Hamblin 1970, MacKenzie 1979) an argument is gradually “rolled off” on request of the dialogue partner. An example would be the following.

### **Example: Tax Relief (I)** (Argument: pmp; pmp $\Rightarrow$ tr)

“Next year, we are going to get a tax-relief, because our politicians promised so.”

P: claim(tr) “I think that we will get a tax-relief.”

O: why tr “Why do you think so?”

P: because pmp  $\Rightarrow$  tr “Because our politicians promised so.”

O: concede tr “OK, sounds convincing.”

### 3. Examples of Socratic-style dialogue

Many of today's generation of dialogue systems regard a (persuasion) dialogue a process in which a claim is provided by one of the dialogue partners, after which an iterative search for the justification of this claim starts. The main claim is successfully defended if this iterative process ends with the opponent (O) conceding the main claim. An illustration is the Tax Relief (I) example above.

Nevertheless, in actual human-style conversations one can also observe a second form of dialogue, which has until now received relatively little attention in formal dialogue. The idea is that instead of asking for the reasons for a specific claim (why/because) one confronts the discussion partner with what appear to be the logical consequences of his own point of view. When these consequences approach absurdity, it is clear that the main claim cannot hold, and the proponent has lost the dialogue. An example of this would be the Tax Relief (II) example below.

#### **Example: Tax Relief (II)**

“Next year, we are going to get a tax-relief”

“But in the current situation, you can only get a tax-relief by accepting a significant budget deficit, which means we will also get a huge fine from Brussels. There goes our tax relief...”

P: claim tr “Next year, we're going to get a durable tax-relief.”

O: but-then tr  $\Rightarrow$  bd “But isn't it the case that, under current economic circumstances, a tax-relief produces a budget deficit?”

P: concede bd “Yes, I think so.”

O: but-then bd  $\Rightarrow$  fb “But you know that a budget deficit results in a fine from Brussels (EU), don't you?”

P: concede fb “Unfortunately, yes.”

O: but-then fb  $\Rightarrow$   $\neg$ tr “Then there wouldn't be any space left for a truly durable tax relief”

P: concede  $\neg$ tr “Oops, you're right.”

In the Tax Relief (II) example party P puts forward the main claim that there will be a durable tax-relief. In a traditional dialogue system, such as MacKenzie's DC or Hamblin's H, player B could then either concede this claim or ask for a reason for it. This, however, is not what happens here. Instead of asking for a justification, player O confronts P with what O sees as the consequence of P's position. After two additional similar steps are made by O, player P is in the position that he has refuted himself.

P's refutation is not unavoidable. The trick is that in some point of the dialogue, P needs a way to avoid being forced to accept the statements that O wants P to accept. An example of how such could be done is shown in the following dialogue.

#### **Example: Tax Relief (III)**

“Next year, we are going to get a tax-relief”

“But in the current situation, you can only get a tax-relief by accepting a significant budget deficit, which means we will also get a huge fine from Brussels. There goes our tax relief...”

P: “France and Germany also have a budget deficit, so the rule that countries with a budget deficit will be fined is probably to be abolished.”

P: claim tr “Next year, we’re going to get a durable tax-relief.”

O: but-then tr  $\Rightarrow$  bd “But isn’t it the case that, under current economic circumstances, a tax-relief produces a budget deficit?”

P: concede bd “Yes, I think so.”

O: but-then bd  $\Rightarrow$  fb “But you know that a budget deficit results in a fine from Brussels, don’t you?”

P: claim  $\neg$ [bd  $\Rightarrow$  fb] “I don’t think that a budget deficit necessarily leads to a fine.”

O: why  $\neg$ [bd  $\Rightarrow$  fb] “Why not?”

P: because bd(F), bf(G)  $\Rightarrow$   $\neg$ [bd  $\Rightarrow$  fb] “Because France and Germany also have budget deficits, and it is unlikely that any EU-rule will be enforced when it is against the national interest of these dominant EU members.”

O: concede tr “OK, now I really think we are going to get a durable tax relief.”

In the example Tax Relief (III) party P successfully produces an undercutter of the principle that a budget deficit leads to a fine. O’s strategy of step by step forcing P to commit himself to an inconsistency has therefore been broken.

In a traditional Socratic dialogue, the aim of the opponent is to have the proponent deny one of proponent’s earlier statements. One could, however, also imagine a weaker form of refutation, in which the proponent is forced to undercut one of the rules he put forward earlier. An example would be Pollock’s Pink Elephant example (Pollock 1995, p. 120, Caminada 2004, p. 136).

## 4. Analysis

It is interesting to compare the use of the “because” dialogue move with the use of the “but then” dialogue move. The “because” dialogue move is used to (iteratively) give reasons for particular statements. A typical dialogue may look as follows:

A: claim  $\Phi$

B: why  $\Phi$

A: because  $\psi \Rightarrow \Phi$

B: why  $\psi$

A: because  $\chi \Rightarrow \psi$

⋮            ⋮

The ‘but-then’ dialogue move is used to try to force the other party to make commitments towards a specific direction. A typical dialogue may look as follows:

A: claim  $\Phi$

B: but-then  $\Phi \Rightarrow \eta$   
 A: concede  $\eta$   
 B: but-then  $\eta \Rightarrow \mu$   
 A: concede  $\mu$   
 ⋮            ⋮

Overall, one can distinguish the following four differences between the “because” statement and the “but-then” statement (see also the figure below).

1. With the “because” statement, reasoning goes *backwards* (abduction); the player being questioned tries to find reasons to support his thesis. With the “but-then” statement, on the other hand, reasoning goes *forward* (deduction); the player being questioned can be forced to make additional reasoning steps.
2. With the “because” statement, the *proponent* of a thesis (like  $\Phi$  in the figure below) tries to find a path (or tree) from the premises to  $\Phi$  (the opponent's task is then to try to defeat this path). With the “but-then” statement, on the other hand, it is the *opponent* of the thesis that tries to find a path (or tree).
3. The path (or tree) constructed using “because” statements should ultimately originate from statements that are accepted to be *true* (such as premises), whereas the path constructed using “but-then” statements should ultimately lead to statements that are considered *false* (contradictions) or at least problematic (undercutters of rules used to justify  $\Phi$ ).
4. With a successfully constructed “because” path (or tree), both the proponent and opponent ultimately become committed to the propositions on the path, whereas with a successfully constructed “but-then” path (or tree), it is possible that only the proponent ends up being committed to the propositions on the path.



In the above analysis, it may seem that an opponent of  $\Phi$  has two options: either trying to construct a but-then path from  $\Phi$ , or trying to prevent the proponent from successfully constructing an undefeated because path. A mixed strategy, however, is sometimes also possible.

## 5. Commitment in Dialogue

In their book, Walton and Krabbe (1995) specify how dialogues can be seen from the perspective of *commitments*. A (propositional) commitment can be regarded as a party's “official” point of view, a public statement of what the party holds to be true and is willing to defend if being questioned.

As an example of how commitments come into existence during the course of a dialogue, consider

again the Tax Relief (I) example.

P:	claim(tr)	$C_P(\text{tr})$
O:	why tr	[no commitments]
P:	because pmp $\Rightarrow$ tr	$C_P(\text{tr}, \text{pmp})$
O:	concede tr	$C_O(\text{tr})$

Here, with the statements “claim”, “because” and “concede”, new commitments come into existence. A mere question (“why”), however, does not directly create any new commitment.

An interesting question is how the “but-then” statements should be seen in terms of commitments. Consider the following statement: “but-then  $\Phi \Rightarrow \eta$ ”. An intuitive reading would be something like: “But from your position  $\Phi$ , doesn't it follow that  $\eta$ ?” or even “If you hold that  $\Phi$  then you actually also hold that  $\eta$ , don't you?”

Although both interpretations are formulated as a question, it is actually not so much information that the party uttering a “but-then” statement is interested in. When one asks a question, one is usually assuming that the other party has some information that one is interested in. In the dialogues of Socrates, at the other hand, it is hard to maintain that Socrates asked his questions because he was genuinely in pieces of knowledge he did not have. Many of Socrates's utterances are merely *disguised* as questions. Classifying them as if they actually are real questions would not do justice to their fundamental nature.

If the “but-then” statement is not to be classified as a question then what should be its classification instead? One possible approach would be to regard the “but-then” statement as a particular form of claim. This, however, leaves open the question of what it is that is claimed. If party O utters something like “But from your position  $\Phi$  doesn't it follow that  $\eta$ ?”, then what is the object of the claim? Surely, O does not necessarily hold that  $\eta$ , just like Socrates was clever enough not to commit himself to the statements that he let others to commit to, or a critical TV-interviewer is not necessarily committed to the statements he manages to extract from the interviewee.

Is it then perhaps the case that an utterance like “But from your position  $\Phi$  doesn't it follow that  $\eta$ ?” creates a commitment on behalf of the hearer? That is, does O's statement “but-then  $\Phi \Rightarrow \eta$ ” create  $C_P(\eta)$ ? Apart from it being odd that an utterance of one party can create a commitment for the other party, there also exist other reasons for rejecting this view. In example Tax relief (III) it is illustrated that a party can sometimes avoid being committed by arguing against the applicability of a rule (like  $bd \Rightarrow fb$ ). Therefore, a move of O of the form “but-then  $\Phi \Rightarrow \eta$ ” does not directly lead to a commitment of the form  $C_P(\eta)$ .

Perhaps the best way to view O's statement “but-then  $\Phi \Rightarrow \eta$ ” in terms of commitments would be to literally take into account the interpretation “If you hold that  $\Phi$ , then you actually also hold that  $\eta$ .” In this way, what O is actually is committed to is *that P is committed to*  $\eta$ : “I hold that you

hold that  $\Phi$ , since this is a (defeasible) consequence of your own explicit commitment  $\Phi$ .” Thus, we have that O's statement “but-then  $\Phi \Rightarrow \eta$ ” creates  $C_O(C_P(\eta))$ .

One interesting way to see commitments has been proposed by Walton and Krabbe (1995). The idea is to regard a (propositional) commitment as that which has to be defended during the course of a dialogue, that which a party is trying to establish. In a Socratic dialogue the questioner (Socrates) ultimately wants his discussion partner to commit himself to a contradiction. In the Tax Relief examples (II and III) the opponent also tries to commit the proponent to a contradiction. This is done in small steps where with each statement of the form “but-then  $\Phi \Rightarrow \eta$ ” O tries to commit P to  $\eta$ . P could then try to argue against the applicability of  $\Phi \Rightarrow \eta$  (like is done in Tax Relief III) and it is O's task not to let this succeed. Thus, what O tries to defend is  $C_P(\eta)$ . From the perspective of commitment in dialogue, this means that  $C_O(C_P(\eta))$ .

It is interesting to re-examine the Tax Relief (III) example in terms of commitments.

P:	claim tr	$C_P(\text{tr})$
O:	but-then tr $\Rightarrow$ bd	$C_O(C_P(\text{bd}))$
P:	concede bd	$C_P(\text{tr}, \text{bd})$
O:	but-then bd $\Rightarrow$ fb	$C_O(C_P(\text{fb}))$
P:	claim $\neg[\text{bd} \Rightarrow \text{fb}]$	$C_P(\text{tr}, \text{bd}, \neg[\text{bd} \Rightarrow \text{fb}])$
O:	why $\neg[\text{bd} \Rightarrow \text{fb}]$	[unchanged]
P:	because $\text{bd}(\text{F}), \text{bd}(\text{G}) \Rightarrow \neg[\text{bd} \Rightarrow \text{fb}]$	$C_P(\text{tr}, \text{bd}, \neg[\text{bd} \Rightarrow \text{fb}], \text{bd}(\text{F}), \text{bd}(\text{G}))$
O:	concede tr, retract $C_P(\text{fb})$	$C_O(\text{tr})$

In the above dialogue, we have omitted  $C_O(C_P(\eta))$  when it already holds that  $C_P(\eta)$ . In general, there exist various relations between nested and unnested commitments. An overview can be found at page 50-52 of (Caminada 2004).

## 6. Socratic-style dialogue and MacKenzie's DC

Some people may argue that for Socratic-style dialogue the “but-then” statement is not strictly necessary and that the associated dialogues could also be modelled in MacKenzie's DC by means of the “resolve” statement. As a side effect there would also be no need for nested commitments. In this way, the following dialogue fragment

P:	claim p	$C_P(\text{p})$
O:	but-then p $\Rightarrow$ q	$C_O(C_P(\text{q}))$
P:	concede q	$C_P(\text{q})$

could also be modelled by

P:	claim p	$C_P(\text{p})$
----	---------	-----------------

O:	concede p	$C_o(p, p \supset q)$
P:	concede $p \supset q$	$C_p(p, p \supset q)$
O:	claim q	$C_o(p, p \supset q, q)$
P:	why q	[unchanged]
O:	resolve “If $p \wedge (p \supset q)$ then q”	[unchanged]
P:	concede q	$C_p(p, p \supset q, q)$

A central role in the second dialogue above is played by MacKenzie's “resolve” statement. With this, one can confront a party with the consequences of his own commitments. If one party asks “why q” and is committed to p as well as to  $p \supset q$  then the other player can utter “resolve p,  $p \supset q$ ”. What makes the above examples different, however, is the fact that the rule referred to in a “but-then” statement can be defeasible. If party O confronts party P with the principle that a budget deficit leads to a fine from Brussels, then player O may find a way out by providing an undercutter of this principle, thus avoiding to endorse that there will be a fine as well. It is this defeasibility that is lacking in the “resolve” statement.

Another important difference between the above two approaches is related to the issue of burden of proof. With the statement “claim  $p \supset q$ ” it is the *speaker* who has to argue to defend this claim, whereas with the statement “but-then  $p \Rightarrow q$ ” it is the *hearer* who would have to argue against the applicability of  $p \Rightarrow q$ , to avoid becoming committed to q.

Some may argue that with the statement “but-then  $p \Rightarrow q$ ” a new rule is introduced which also has to be defended. In many occasions, however, the rule “ $p \Rightarrow q$ ” may be a piece of information that both dialogue partners adhere to as an effect of of being members of a particular group. For instance, two lawyers arguing about a particular case are both bound to the general principles in the relevant jurisdiction. More generally, there are various principles that appear strong enough to have the burden of proof on those who want to deny the validity or applicability. In these cases, the “but-then” statement would be most appropriate.

## References

- Caminada, M.W.A. (2004). *For the Sake of the Argument; explorations into argument-based reasoning*. Ph.D. thesis Vrije Universiteit Amsterdam.
- Fulda, J.S. (2000). The logic of “improper cross”. *Artificial Intelligence and Law* 8, 227-341.
- Hamblin, C.L. (1970). *Fallacies*. Richard Clay (The Chaucer press) Ltd., Bungay, Suffolk.
- MacKenzie, J.D. (1979). Question-begging in non-cumulative systems. *Journal of Philosophical Logic* 8, 117-133.
- Milward, D & M.A. Beveridge (2004). Ontologies and the structure of dialogue. *Proceedings of CATALOG, 8<sup>th</sup> Workshop on the Semantics and Pragmatics of Dialogue*. Barcelona, Spain.
- Parsons, S & M. Wooldridge & L. Amgoud (2003). Properties and complexity of formal inter-



- agent dialogues. *Journal of Logic and Computation* 13(3), 247-376.
- Pollock, J.L. (1992). How to reason defeasibly. *Artificial Intelligence* 57, 1-42
- Pollock, J.L. (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person*. Cambridge, Massachusetts: The MIT Press.
- Prakken, H. & G. Sartor (1997). Argument-based reasoning with defeasible priorities. *Journal of Applied Non-classical Logics* 7, 25-75.
- Prakken, H. (2005). Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation* 15, 1009-1040.
- Skidmore, J. (2002). Skepticism about practical reasoning: transcendental arguments and their limits. *Philosophical Studies* 109, 121-141.
- Walton, D.N. & E.C.W. Krabbe (1995) *Commitment in Dialogue, Basic Concepts of Interpersonal Reasoning*. State University of New York Press.