

Part A Chapter 3

Motivation-Based Behaviour

The third class of behavioural patterns to be discussed is motivation-based behaviour. In this case an agent does not simply respond to stimuli, neither immediately, nor to an internally maintained world state model based on stimuli in the past. In addition to the observations and world state model, assumed *motivational attitudes* play an important role in determining its actions. These motivational attitudes can be based on the agent's own character (for example, an agent's character may be that it always wants to eat), but also on specific aspects of the agent's own state (such as being hungry), that have a more dynamic character.

1 External Dynamics Characterizing Motivation-Based Behaviour

As in Chapter 2, first we take the external viewpoint. Assume that the external traces depicted in Table 1 are observed. The same questions are addressed as before:

- What is the pattern behind this observed behaviour?
- Which external dynamic properties characterize the pattern of these traces?
- Which assumed internal state properties realize this externally observed behaviour?
- What is the pattern of dynamics of these internal state properties?
- How can these internal dynamics be characterized by dynamic properties?

Compared to delayed response behaviour, the puzzle to find answers to these questions is a bit harder this time.

<i>time trace</i>	<i>time point 0</i>	<i>time point 1</i>	<i>time point 2</i>	<i>time point 3</i>	<i>time point 4</i>	<i>time point 5</i>
<i>trace 1</i>	food screen	no food no screen	food screen	food screen	food no screen	food no screen goto p2
<i>trace 2</i>	no food no screen	food no screen	no food screen	no food no screen	no food no screen	no food no screen
<i>trace 3</i>	no food no screen	no food no screen	food screen	food no screen	food no screen	food no screen
<i>trace 4</i>	food no screen	no food no screen	food screen	food screen	food screen	food screen
<i>trace 5</i>	no food no screen	food no screen	no food screen	food screen	food no screen	food no screen goto p2
<i>trace 6</i>	food no screen	no food no screen	food screen	food screen	food no screen	food no screen goto p2
<i>trace 7</i>	food screen	food no screen	food screen	food screen	food no screen	food no screen

Table 1 Example set of observed traces of motivation-based behaviour

Answers to the first two of these questions are addressed (from a *behaviourist perspective*) in this section. The other three are addressed (from a *functionalist perspective*) in Section 2, and (from two variants of this perspective: the *instrumentalist perspective* and the *realist perspective*) in Section 3. The following informally expressed dynamic property is satisfied by the traces in Table 1:

The agent will go to p2 at each time point that it observes the absence of the screen and in the past the agent observed that food at p2 disappeared, after which again food was present.

In structured semiformal form:

EMB1

at any time point t1
 if the agent observes the absence of the screen,
 and there exists time points s1, s2, and s3 with $s1 < s2 < s3 \leq t1$ such that
 at s3 the agent observed the presence of food at p2,
 and at s2 the agent observed the absence of food at p2,
 and at s1 the agent observed the presence of food at p2,
 then after time point t1 the agent will go to p2

This type of behaviour is slightly more complex than the previous types addressed in Chapter 2. Also here the question can be posed what increased complexity of the environment can be coped with. Possibly answers to this question can be found in environments where food can be absent for longer time periods, so that food-hoarding may be important to survive.

Explanation of Motivation-Based Behaviour from a Behaviourist Perspective

Given the above description from an external, behaviourist perspective the following question is addressed.

Ø How does a behaviourist explanation of motivation-based behaviour look like?

The following example explanation illustrates this.

Why does the animal go to p2 ?

The animal goes to p2, because it just observed the absence of the screen, and at an earlier point in time it observed food at p2, whereas at a still earlier point in time it observed the absence of food at p2, and at yet another still earlier point in time it observed the presence of food at p2.

This explanation gets the form of a kind of historical perspective on the course of events that have lead to the current behaviour. It has only reference to causes in an indirect manner (except the observation of the absence of the screen).

Interesting variants of this motivation-based behaviour, in combination with delayed response behaviour occur if also cups hiding the food play a role. For example, in the traces in Table 1, after time point 3 a cup can be placed at p2, hiding the food, if it was present. Since it is assumed that the animal also performs delayed response behaviour, the outcomes will be the same. In this case an interesting difference can be noticed between traces 1 and 7. If the animal has a world state model, then this will be the same for both cases. However, the animal does not act the same. Therefore its behaviour is not just dependent on its world model and direct observations, i.e., it does not perform reactive behaviour. In particular, the behaviour depends on internal maintenance of aspects of history that are beyond a world state model. In Section 2 these aspects of history, and their internal maintenance will be addressed in more detail.

2 Internal Dynamics Generating Motivation-Based Behaviour

Why would the behavioural pattern identified above be in the class of motivation-based behaviour? In this subsection this question is addressed by assuming two internal state properties: *desire* and *intention*. Moreover, also a belief is relevant: an *additional reason* (to generate an intention from a desire), and an observation of an *opportunity* (to actually perform an intended action). The interdependencies between these state properties, depicted by arrows in Figure 1 are interpreted as dependencies between these state properties in the process of generating them.

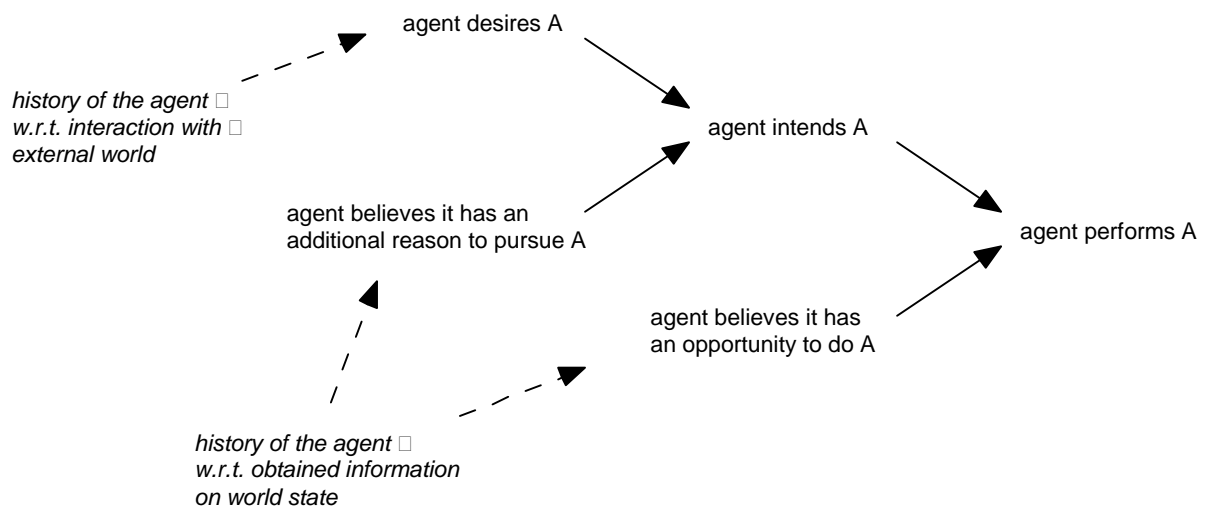


Figure 1 Relationships between the motivational notions

Let's assume that internal properties $b1$, d and i are available with the following dynamics (partly taken from the figure):

IMB1 Desire Generation

at any point in time
 if the agent observes the absence of food at $p2$,
 and there exists an earlier time point at which
 the agent observed the presence of food at $p2$,
 then the internal state property d will hold

IMB2 Intention Generation

at any point in time
 if the internal state property d holds,
 and the internal state property $b1$ holds
 then the internal state property i will hold

IMB3 Action Generation

at any point in time
 if the internal state property i holds,
 and the agent observes the absence of the screen
 then the agent will go to $p2$

In addition, the following dynamic properties of $b1$ are taken over from Section 2.

IDR1 Belief Generation

for all time points
 if the agent observes that food is present at position $p2$,
 then internal state property $b1$ will hold

IDR4 Belief Persistence

for all time points $t1$ and $t2$ with $t1 < t2$
 if internal state property $b1$ holds at $t1$,
 and between $t1$ and $t2$ the agent does not observe that food is not present at position $p2$,
 then internal state property $b1$ holds at $t2$

From these properties **IMB1** is not in the executable format as defined in Chapter 2. To replace the property by executable properties, the additional internal property p is assumed (a kind of pre-desire), with the following dynamics:

IMB1a Predesire Generation

at any point in time
 if the agent observes the presence of food at p2,
 then the internal state property p will hold

IMB1b From Predesire to Desire

at any point in time
 if the agent observes the absence of food at p2,
 and the internal state property p holds,
 then the internal state property d will hold

Notice that these properties together entail **IMB1**. The internal state properties p, d and i are assumed to persist, for convenience this persistence is assumed unconditional:

IMB4 Desire Persistence

for all time points
 if internal state property d holds,
 then for every later time point internal state property d holds

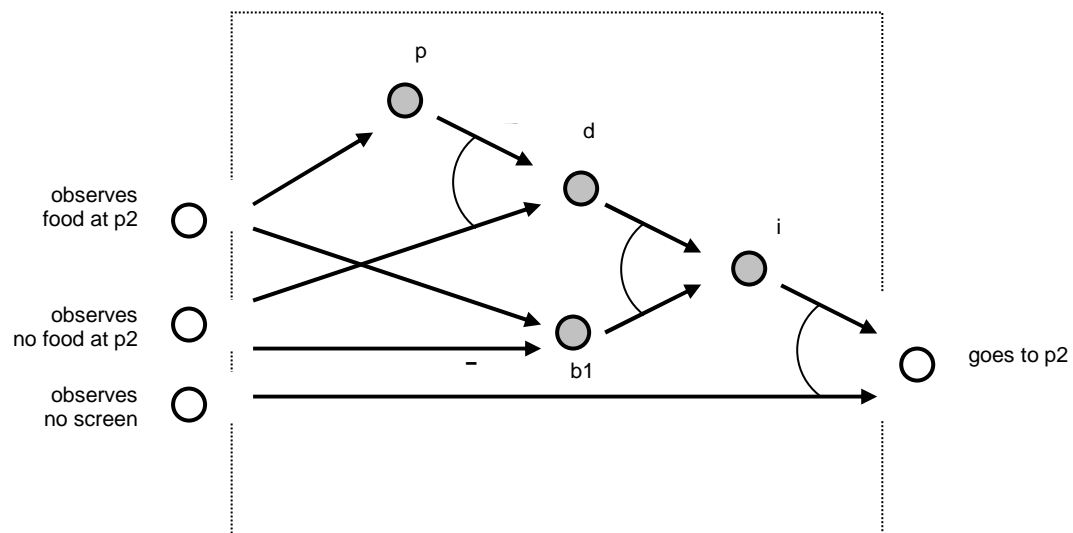
IMB5 Intention Persistence

for all time points
 if internal state property i holds,
 then for every later time point internal state property i holds

IMB6 Predesire Persistence

for all time points
 if internal state property p holds,
 then for every later time point internal state property p holds

The dynamic properties **IMB1a**, **IMB1b**, **IMB2**, **IMB3**, **IMB4**, **IMB5**, **IMB6**, **IDR1**, **IDR4** defining the internal dynamics are depicted in a graphical form as follows.



The internal dynamic properties expressed in this way specify the functional roles of the internal state properties *p*, *d*, *i*, and *b1*. Together these dynamic properties entail the external dynamic property **EMB1**, as is illustrated in the following internal traces. The first one illustrates realisation of external trace 1. Here all internal properties are generated, and also the opportunity is observed. Hence the action goto *p2* is generated.

	<i>time point 0</i>	<i>time point 1</i>	<i>time point 2</i>	<i>time point 3</i>	<i>time point 4</i>	<i>time point 5</i>
<i>time trace 1</i>						
<i>input</i>	food screen	no food no screen	food screen	food screen	food no screen	food no screen
<i>internal</i>		<i>p</i> <i>b1</i>	<i>p</i> <i>d</i>	<i>p</i> <i>d</i> <i>b1</i>	<i>p</i> <i>d</i> <i>i</i> <i>b1</i>	<i>p</i> <i>d</i> <i>i</i> <i>b1</i>
<i>output</i>						goto <i>p2</i>

The second internal trace, which illustrates realisation of external trace 2 is the following. Here the desire is generated, but the intention is not generated, and therefore no goto *p2* action occurs.

	<i>time point 0</i>	<i>time point 1</i>	<i>time point 2</i>	<i>time point 3</i>	<i>time point 4</i>
<i>time trace 2</i>					
<i>input</i>	no food no screen	food no screen	no food screen	no food no screen	no food no screen
<i>internal</i>			<i>p</i> <i>b1</i>	<i>p</i> <i>d</i>	<i>p</i> <i>d</i>
<i>output</i>					

Yet another internal trace, which illustrates realisation of external trace 3 is the following. Here the desire is never generated, hence no intention is generated and no action goto *p2* occurs.

	<i>time point 0</i>	<i>time point 1</i>	<i>time point 2</i>	<i>time point 3</i>	<i>time point 4</i>
<i>time trace 3</i>					
<i>input</i>	no food no screen	no food no screen	food screen	food no screen	food no screen
<i>internal</i>				<i>p</i> <i>b1</i>	<i>p</i> <i>b1</i>
<i>output</i>					

The following internal trace illustrates realisation of external trace 4. Here all internal properties are generated. However, the opportunity does not occur, and therefore the intended action is not performed.

	<i>time point 0</i>	<i>time point 1</i>	<i>time point 2</i>	<i>time point 3</i>	<i>time point 4</i>	<i>time point 5</i>
<i>time trace 4</i>						
<i>input</i>	food no screen	no food no screen	food screen	food screen	food screen	food screen
<i>internal</i>		p b1	p d	p d i b1	p d i b1	p d i b1
<i>output</i>						

Notice that behaviour may occur that may be considered as unreasonable, for example, in the following trace:

	<i>time point 0</i>	<i>time point 1</i>	<i>time point 2</i>	<i>time point 3</i>	<i>time point 4</i>	<i>time point 5</i>	<i>time point 6</i>
<i>time trace 5</i>							
<i>input</i>	food screen	no food no screen	food screen	food screen	food screen	no food no screen	no food no screen
<i>internal</i>		p b1	p d	p d i b1	p d i b1	p d i b1	p d i
<i>output</i>							goto p2

This trace is a possible trace that can occur, but, more than that, given the dynamics of the external world, actually is forced to occur by the dynamic properties expressed above. In particular, the trace shows that the absolute persistence of internal property *i* as expressed in **IMB5** is too much to be reasonable. Actually this property just can be left out, to avoid this undesirable behaviour. If property **IMB5** is left out, behaviour given by the following trace becomes possible:

	<i>time point 0</i>	<i>time point 1</i>	<i>time point 2</i>	<i>time point 3</i>	<i>time point 4</i>	<i>time point 5</i>	<i>time point 6</i>	<i>time point 7</i>	<i>time point 8</i>
<i>time trace 1</i>									
<i>input</i>	food screen	no food no screen	food screen	food screen	food screen	no food no screen	no food no screen	no food no screen	no food no screen
<i>internal</i>		p b1	p d	p d b1	p d i b1	p d i b1	p d	p d	p d
<i>output</i>							goto p2		

Now *i* does not have to be persistent anymore, but still *d* is persistent. This may be considered a bit strong for what is meant to be an example of motivation-based behaviour; since something inside has changed forever, it looks more like adaptive behaviour, and as such would better fit in Chapter 4: after one experience (of probably extreme hunger), the agent is desiring food forever. A more modest form of motivation-based behaviour can be found when also the action of eating is taken into account, and after eating the desire disappears. Then the persistence property for *d* should be different:

IMB4'

for all time points *t*₁ and *t*₂
 if internal property *d* holds at *t*₁,
 and between *t*₁ and *t*₂ no eat action occurred
 then internal property *d* holds at *t*₂

With **IMB4** replaced by **IMB4'** and **IMB5** left out a more typical case of motivation-based behaviour is obtained. In this case all internal properties that hold at some point in time can change back to not holding at some future time point.

Explanation of Motivation-Based Behaviour from a Functionalist Perspective

Given the above description from an internal, functionalist perspective the following question is addressed.

Ø What does a functionalist explanation of motivation-based behaviour look like ?

The following example explanation illustrates this.

Why does the animal go to *p*₂ ?

The animal goes to *p*₂, because it just observed the absence of the screen, and it had the intention to go to *p*₂.

Why did the animal have this intention to go to *p*₂ ?

The animal had the intention to go to *p*₂, because it had the desire to go to *p*₂ and the belief that there was food at *p*₂.

Why did the animal have the desire to go to *p*₂; and why did it believe that food was present at *p*₂ ?

The animal desired to go to *p*₂, because it had this desire persisting for a longer time already. The same holds for the belief that food was present at *p*₂.

But then, why did these desire and belief start to be there ?

The desire started to be there because the animal observed that there was no food at *p*₂, and the animal had a predesire.

The belief that food is present at *p*₂ started to be there, because the animal observed the presence of food at *p*₂.

Why did the animal have this predesire ?

The animal had this predesire, because it had this predesire persisting for a longer time already.

Why did this predesire start to be there ?

The predesire started to be there because the animal observed the presence of food at p2.

Also this explanation gets the form of a kind of historical perspective on the course of events that have lead to the current behaviour. However, in contrast to the explanation in Section 1, in this case this history can be partitioned according to executable steps. Therefore this explanation has only reference to causes in a more direct manner.

Further questions

A number of further questions can be addressed:

- Ø Which external property (instead of **EMB1**) would describe the behaviour for the case of having conditional persistence property **IMB4**?
- Ø What do the properties look like if response times are taken into account ?
- Ø How can other variants of motivation-based behaviour be analysed; for example, desires that depend on hunger, which occurs after some period without eating, and disappears after eating ?

3 The Attribution of Motivations

Explanations of behaviour from a behaviourist perspective make use of the input the organism receives (stimuli) and the generated output (behaviour) and their relationships, without involving any internal state properties. Explanations of behaviour from a functionalist perspective, in addition, take internal state properties into account, and their mutual relationships: *functional roles*. Note that if the internal dynamics are characterised by a set of dynamic properties, then the functional role of an internal state property p is described by the *subset of those dynamic properties in which p occurs*.

The status of the internal state properties involved in a functionalist explanation can be of two types:

(a) *Realist perspective*

The internal state properties are (known to be) related to certain ‘real’ physical (e.g., neurological) state properties; these physical state properties are called *realisers* of the internal state properties

(b) *Instrumentalist perspective*

The internal state properties are not (known to be) related to ‘real’ physical state properties: no realisers are assumed. The internal state properties are an instrument, for example, to do predictions on behaviour, but do not necessarily exist in reality

The first, realist perspective allows to provide a causal basis of an explanation. The *functional roles* of the internal state properties correspond to *causal relationships* between their physical realisers. An explanation based on the internal state properties can be replaced by a causal explanation in terms of the realisers and their causal relationships. For example, an agent damaging its tissue feels pain and subsequently will move away. If having pain can be related to a physical state *n*, then the explanation ‘moving away because of pain because of tissue damage’ can be replaced by ‘moving away because of physical state *n* because of tissue damage’, where the ‘because of’ relations in the latter case are based on causal relationships between physical state properties.

In the second, instrumentalist perspective no realisers of the internal state properties are assumed. As an instance of the *instrumentalist perspective* and opposed to explanations from a direct physical perspective (the physical stance), in (Dennett, 1987, 1991) the *intentional stance* (or folk-psychological stance) is put forward. In (Dennett, 1987), pp. 37-39, he explains the advantage of intentional stance explanations for mental phenomena over physical stance explanations:

‘Predicting that someone will duck if you throw a brick at him is easy from the folk-psychological stance; it is and will always be intractable if you have to trace the protons from brick to eyeball, the neurotransmitters from optic nerve to motor nerve, and so forth.’ (Dennett, 1991), p. 42

To function more efficiently in an organisation, it is useful if agents fulfilling a certain role in the organisation can reason about the possible behaviour of the agents in other roles, for example using the intentional stance. For example, to an agent functioning within an organisation it may be very helpful to have capabilities to predict in which circumstances certain inappropriate desires or intentions are likely to arise as a basis for the behaviour of a colleague within the organisation, either to avoid the arising of these intentions by preventing the occurrence of circumstances that are likely to lead to them, or if these circumstances cannot be avoided, by anticipating consequences of the intentions. Such capabilities of *anticipatory reasoning* about the behaviour of colleagues in an organisation are quite important for an organisation to function smoothly.

According to the intentional stance, an agent is assumed to decide to act and communicate based on intentional notions such as beliefs about its environment and its desires and intentions. These decisions, and the intentional notions by which they can be explained and predicted, generally depend on circumstances in the environment, and, in particular, on the information on these circumstances just acquired by interaction (i.e., by observation and communication), but also on information acquired by interaction in the past. To be able to analyse the occurrence of intentional notions in the behaviour of an observed agent, the observable behavioural patterns over time form an empirical basis; cf. (Dennett, 1991).

Below, the dependencies assumed between the different intentional notions are assumed to be as in Section 2, with the difference that this time no internal properties within the observed agent are assumed. Instead, the notions are attributed by an external observer.

Cases in Organisations

Viewed from a dynamic perspective, organisational structure provides specifications of constraints on the dynamics of role behaviour and interactions. By these specifications to a certain extent coordinated dynamics is enforced to the organisation. In human organisations role specifications usually do not completely prescribe behaviours, however. To a greater or lesser extent some space of freedom in behaviour and personal initiative is allowed. This freedom has its positive elements; in the first place, human agents can find more satisfaction and do their work with higher quality if they can do things in their own way. In the second place an organisational structure does not anticipate on all possible circumstances; in unforeseen situations it can be beneficial if agents have some space to improvise.

The reverse of the medal, however, is that this freedom also may provide possibilities to agents to show certain behaviours (based on their individual characteristics and interests) that decrease the extent of coordination. To function more efficient in an organisation, where roles do not completely prescribe behaviour, it is useful if agents fulfilling a certain role in the organisation can reason in an anticipatory sense about the behaviour of the agents in other roles, for example, using the intentional stance. This section addresses some examples of the phenomena described for human organisations as an illustration.

- (a) An employee has done something very important very wrong, and deliberates whether or not to tell his manager:
'If he believes that I am the cause of the problems, he will try to fire me.'
- (b) An employee has encountered a recurring problem, and knows a solution for this problem, on which he would like to work. He deliberates about how to propose to his manager this solution.
'If I tell this solution immediately he will not believe that the problem is worth working on it. If I make him aware of the problem, and do not tell a solution, he only will start to think himself about it for a while, without finding a solution, and then forget about it. If I make him aware of the problem and give some hints that direct him to a (my) solution, he will believe he contributed to a solution himself and want me to work on it.'
- (c) A manager observes that a specific employee in the majority of cases functions quite cooperatively, but shows avoidance behaviour in other cases. In these latter cases, the employee starts trying to reject the task if he believes that his agenda already was full-booked for the short term, it is not clear to him whether somebody else is not capable of doing the task, and he believes colleagues are available with less full-booked agendas.

Further observation by the manager reveals the pattern that the employee shows avoidance behaviour, in particular, in cases that a task is only asked shortly before its deadline, without the possibility to anticipate on the possibility of having the task allocated. The manager deliberates about this as follows:

'If I know beforehand the possibility that a last-minute task will occur, I can tell him the possibility in advance, and in addition point out that I need his unique expertise for the task, in order to avoid the behaviour that he tries to avoid the task when it actually comes up.'

The reasoning processes on predicted behaviours described in the examples (a) to (c) can be based on prescribed role behaviours (as may be the case in (a)), or on an analysis of the other agent's personal motivations (as is the case in (b) and (c)). To illustrate this, example (c) is addressed by making the following interpretation.

The *desire* to avoid a task is created after time t by the employee if the following holds for the history:

- at time t the employee heard the request to perform the task
- at time t the employee observes that the task has to be finished soon
- the employee did not hear of the possibility of the task at any earlier time point

The *intention* to avoid a task is generated after time t if the following holds for the history:

- the desire to avoid the task is available at time t
- the belief that colleagues are capable of doing the task is available at time t
- the belief that colleagues are not full-booked is available at time t

The *action* to avoid the task is generated at time t if the following holds for the history:

- the intention to avoid the task is available at time t
- the belief that the employee's own agenda is full-booked is available at time t

Given this interpretation it can be illustrated how the manager agent can reason and act in an anticipatory manner to avoid the employee's avoidance desire, intention and/or action to occur. This can be done in the following three manners:

Avoiding the desire to occur

This can be obtained by communicating in advance to the employee that possibly a last minute task will occur. This would make the third condition in the definition of the temporal desire statement fail.

Avoiding the intention to occur (given that the desire occurs)

This can be obtained by refutation of the reason to generate the intention, e.g., by telling the employee that he is the only one with the required expertise.

Avoiding the action to occur (given that the intention occurs)

This can be obtained by refutation of the opportunity, e.g., by taking one of the (perhaps less interesting) tasks from his agenda and re-allocating it to a colleague.

The following questions can be addressed to elaborate further on this case.

- Ø Can you relate the analysis given above to the dynamic properties of beliefs, desires and intentions identified in Section 2?
- Ø Describe a typical behaviour of the employee in case (c). Explain this behaviour from an instrumentalist perspective based on attributed intentional notions.
- Ø Why would such an explanation be valid? Which criteria should be satisfied in order to obtain a valid explanation?
- Ø Address cases (a) and (b) in a similar manner.

References

- Dennett, D.C. (1987). *The Intentional Stance*. MIT Press. Cambridge Mass.
- Dennett, D.C. (1991). Real Patterns. *The Journal of Philosophy*, vol. 88, pp. 27-51.