Foundations for an Ontology of Belief, Desire and Intention

Fumiaki TOYOSHIMA [a,1], Adrien BARTON [b,a,2] and Olivier GRENIER [c]

[a] *GRIIS, Université de Sherbrooke, Quebec, Canada*
[b] *Institut de Recherche en Informatique de Toulouse, CNRS, Toulouse, France*
[c] *CIRST, Université du Québec à Montréal, Quebec, Canada*

Abstract. Belief, desire, and intention are central notions in mentality and agency. We provide conceptual and formal foundations for an ontology of those mental entities. In this framework, beliefs and desires have a dual face: dispositional and occurrent. As distinct from beliefs and desires, intentions are dispositions to actions that emerge from a decision process in which occurrent beliefs and occurrent desires interact. We also discuss how our theory can be extended to some major philosophical accounts of desires, and cognitive biases such as wishful thinking.

Keywords. Belief, desire, intention, disposition, occurrent

1. Introduction

Belief, desire, and intention (which we will call "BDI entities") merit careful investigation because they are of paramount importance for an ontology of mental reality and agency. Inspired by Bratman's [1] philosophy, for instance, the BDI model of agency recognizes the primacy of the BDI entities in practical reasoning and rational actions [2]. It has been utilized in formal ontology of mind [3] and action [4]. Relatedly, the notion of goal and related entities (e.g. trying) have been formally explored [5], as it plays a vital role in the BDI model as well as in commonsense psychology [6]. To take another example, cognitive processes and representations have been investigated in the Mental Functioning Ontology (MF) [7] (which aims to serve as a mid-level ontology for mental functioning), and religious and spiritual beliefs are formalized along with the MF in the Web Ontology Language (OWL) [8]. Nonetheless, the BDI entities (*inter alia* desire and intention) would tend to be loosely characterized, especially so that their parent types are identified, but sometimes with no further detailed examination. Examples include belief as a "mental disposition" in the MF, the BDI entities as "mental moments" in UFO-C [9] (which is a module for social and intentional entities in the upper ontology the Unified Foundational Ontology), and the BDI entities as "mental states" in a Deontic Cognitive Event Ontology [10] (which provides an OWL support of representation and reasoning on complex cognitive information).

In this paper we aim to pursue an ontological analysis of belief, desire, and intention. Along with previous works on belief [11], affordance [12-14], and directing actions [15], this paper is part of our project to build an ontology that covers the core categories and relations concerning agency, cognition, and actions. It will also give an impetus to develop and integrate existing BDI-based information systems and ontologies. For this purpose, we present previous formal-ontological works on dispositions (Section 2) and provide an ontological analysis of the BDI entities, actions, and their relationship along dispositional lines (Section 3). Then, we provide a core formalization of our ontology of the BDI entities and show its compatibility with some additional axioms or hypotheses (Section 4). We also discuss the relationships of our account with existing philosophical accounts of desires as well as with affordances and mental content (Section 5). Finally, we conclude the paper with some remarks on future work (Section 6).

We will use the following scenario named "(HEATER)" as a driving example, while taking for granted the notions of agent and action (see e.g. [4] for detailed discussion). An agent John is at his apartment. During a winter day, he feels cold and desires to get warm. There is a heater in his apartment. He believes that pushing a button of the heater will activate the heater. He also believes that the activated heater will heat up the apartment. He then intends to push the button of the heater and finally performs the action of pushing the button. Next to first-order logic, we will use OWL (using the Manchester Syntax [16]), which is a representation language for ontologies based on Description Logics. Terms for universals will be italicized and terms for particulars and relations will be written in bold.

2. Methodology: A dispositional approach

To describe mental mechanisms found in (HEATER) and other similar cases, we will employ an ontological distinction between continuants (e.g. objects), which persist through time; and occurrents (e.g. processes), which extend through time while having temporal parts and which typically have as participant some continuant. (Other occurrent-related terms "event" and "state" will not be used for simplicity.) We will also utilize previous formal-ontological works on dispositions [17-19], as dispositions are valuable for modeling of various entities (see [20] for a discussion about the relevance of dispositions to scientific ontologies). Note that preceding works on BDI entities (such as Bratman's [1]) do not necessarily see them as dispositional, but our dispositional approach aims for a "core basis" for an ontology of the BDI entities whose full development may require introducing other more specific categories and relations.

A disposition is a property that is linked (by a relation **realized_in**) to a realization, namely to a specific possible behavior of the bearer (such as an object) of the disposition. To be realized in a process, a disposition needs to be triggered (**has_trigger**) by some other process. Classical examples include fragility (the disposition to break when pressed with a force) and solubility (the disposition to dissolve when put in a solvent). Characteristically, dispositions may exist even if they are not realized or even triggered: for example, a glass is fragile even if it never breaks or even if it never undergoes any shock. We will also focus on "sure-fire dispositions" [17] whose realizations necessarily occur once the disposition has been triggered, as well as "single-track dispositions" [17] which have one kind of realizations and one kind of triggering processes. The term "disposition" will henceforth refer to a sure-fire and single-track disposition unless otherwise stated. (Note that our resulting theory of the BDI entities can extend to other

kinds of dispositions such as "multi-track dispositions" [18,21] which have different kinds of realizations according to different kinds of triggers.)

There are two major frameworks to represent dispositions, called (ONLY) and (PARTHOOD) [18]. The (ONLY) framework (first developed by Röhl and Jansen [17]) characterizes a disposition by pointing to its classes of triggers and realizations: a disposition **d** whose class of realization is *R* and whose class of triggers is *T* would be formalized by: **d realized_in** only *R* and **d has_trigger** only *T*.

The (PARTHOOD) framework [18] considers that it is usually not possible to list the whole class of triggers and realizations of a disposition. For example, the specific shock on a glass would trigger its fragility, but so would the process that extends one millisecond earlier and after. Similarly, the specific breaking process of the glass would be a realization of its fragility, but so would the process that extends further to the glass pieces flying apart. Therefore, it introduces the class *TRmin*(**d**) of minimal triggers of a disposition *d* such that, informally speaking, their instances are the "smallest causal process" which exceeds the threshold value for causation. Formally, it is defined as the class of triggers of *d* for which no proper part is a trigger of *d*:

TRmin(**d**) EquivalentTo [(**trigger_of d** and not (**has_proper_part** o **trigger_of d**)]

where **trigger_of** is an inverse relation of **has_trigger**. It also introduces the class *Rmax*(**d**) of maximal realizations of *d*, which is the class of resulting whole causal chain of processes. Formally, it is defined as the class of realizations of *d* which are not proper parts of another realization of *d*:

Rmax(**d**) EquivalentTo [(**realization_of d** and not (**proper_part_of** o **realization_of d**)]

where **realization_of** is an inverse relation of **realized_in**. Two sure-fire single-track dispositions are then considered as identical if and only if they have the same instance of categorical basis, the same class of minimal triggers, and the same class of maximal realizations. Unlike the (ONLY) model, this (PARTHOOD) model of dispositions avoids "disposition multiplicativism" (that is, the excessive arbitrary proliferation of dispositions) [18].

Finally, a recent mereological theory of dispositions [19] specifies several kinds of parthood relations between dispositions. Among them, the **mod-part_of** relation formalizes several possible pathways, or *mode*s, of realizations of dispositions: e.g. the ferromagnetic disposition of this magnet having two mod-parts, i.e. its disposition to attract another magnet when facing an unlike pole and its disposition to repulse a magnet when facing a like pole.

3. An ontological account of belief, desire, and intention

3.1. Belief

Let us begin by discussing belief.[3] In (HEATER), John believes that pushing a button of a heater will activate the heater. The first thing to note is that John's belief exists even

---

[3] Note that the term "belief" is polysemous. One may sometimes use the term "John's belief" to refer to the truth-evaluable content of John's belief: e.g. "John and James both have the very same false belief that the earth is flat". In philosophy, this content is generally taken to be a proposition [22]. The controversial nature of mental contents or propositions is outside the scope of this paper (but see Section 5.5). We will instead focus

when he is not consciously thinking about it, e.g. when he is sleeping. John's belief exists in virtue of some mentally relevant feature of his cognitive (neutral) system. This consideration would suggest a dispositional characterization of belief, since dispositions are properties that are physically grounded and that do not always need to be realized, i.e. activated. The next question to be addressed is how belief (as a disposition) can be realized. A naive idea [23] is that beliefs are dispositions to perform certain kinds of actions: that is, dispositions to behave in a certain way. John's belief that pushing a button of a heater will activate the heater is realized when he performs the action of activating the heater by pushing its button, for instance.

This approach does not seem to account for the nature of belief, however. Intuitively, John can have a belief to that effect even if he is totally paralytic and he is not able to press a button of the heater. One may counter that beliefs are rather dispositions to act if further conditionalized: e.g. John's belief is a disposition to activate the heater if he is physically capable of pressing its button. This proposal would nevertheless only capture the practical (behavioral) dimension of belief (how belief is related to action), but not how belief is connected to the purely theoretical (cognitive) attitude of taking something to be the case. Imagine an omniscient spirit with no power of action at all: he would have many beliefs about the world, but no disposition to act [11].

We therefore hold that a dispositional belief is not realized by physically performing actions, but by some mental process that we call "occurrent belief": namely, the cognitive process of taking something to be the case [11]. In (HEATER), John may believe that pushing on a button of a heater will activate the heater, but this (dispositional) belief is not continuously realized (or, activated) in his mind, as when the weather is hot. In a cold day, on the other hand, he deliberates whether he should press a button of the heater and his dispositional belief d_{BEL} is realized, at time t_0, in a process o_{BEL} of him taking the pressing of the button of the heater to be responsible for the activation of the heater (formally: d_{BEL} **realized_in** o_{BEL}) – a process that will, as we will see, play a causal role in his decision process. In general, an agent may have a dispositional belief even when asleep or unconscious. Briefly, we suggest the following terms and their characterizations:

- Dispositional belief: A disposition that can be realized in an occurrent belief.
- Occurrent belief: A mental process of taking something to be the case.[4]

One may suspect that those characterizations of dispositional belief and occurrent belief are circular, as "taking something to be the case" is usually taken to be synonymous with "believing". But those characterizations can be taken as *elucidations* of the terms in question. As a matter of fact, upper-level entity terms (e.g. "continuant") are hardly definable without circularity and they can be at best elucidated together with the examples to illustrate the entities to which they can apply (cf. [26, p. 89]). Our elucidations of the term "belief" would thus serve to classify two different entities to which it refers to: a belief as a disposition and a belief as a process. This could be likened to a dispositional account of diseases provided in the Ontology for General Medical

---

on John's belief as an entity in John's mind. That is, even if John and James both believe that the Earth is flat, they do not have the same mental state (in a non-technical sense of the term) of belief. Note that this scope for belief will apply, *mutatis mutandis*, to our later discussion on desire and intention.

[4] Our dual account of belief may have a historical root [24]. For another alternative, one may attempt to posit a single kind of belief, for example, by using the notion of "process as a continuant-like occurrent" [25]. Note that these two different approaches to belief can apply for desire, which we will analyze below.

Science [27]: a disease (e.g. epilepsy) is a subtype of disposition that is manifested by undergoing "pathological processes" (e.g. epileptic seizures).

3.2. Desire

We move onto a discussion on desire. Intuitively, there seems to be an intimate connection between desires and behavioral dispositions because the former are closely related to motivation, which is in turn related to the latter [28]. In (HEATER), John is *prima facie* motivated to do things that he believes will result in his warmth. It is however implausible to think that desire is just a disposition towards an action because desires can be active in the agent's mind without exerting a causal effect on its behavior [29]. John's desire to be warm can be active on a cold day, but still not affect his behavior if it is countered by a greatest desire to save energy.

Based on this observation, we adopt a distinction similar to the one drawn by Schroeder [29] between "standing desires" and "occurrent desires", the former being *potentially* active during its existence, and the latter being *actually* and constantly active during its existence. This dual view of desire leads reasonably to the following "desire counterparts" of dispositional and occurrent beliefs [30]:

- Dispositional desire: A disposition that can be realized in an occurrent desire.
- Occurrent desire: A mental process of wanting something to be the case.

In (HEATER), John's dispositional desire d_{DEL} to get warm is realized in his occurrent desire o_{DEL} to get warm (formally: d_{DEL} **realized_in** o_{DEL}). It should be emphasized that, with the same set of desires, an agent could act in multiple ways depending on her beliefs. For instance, John could have decided to put on a sweater, in virtue of his belief that putting on a sweater will get him warmer. To understand John's action of pushing a button of the heater therefore requires considering not only his occurrent desire o_{DEL} but also its interaction with o_{BEL} and his other occurrent belief that the activated heater will warm him up effectively. We will detail this point later.

3.3. Intentions

3.3.1. Intentions and dispositions

We will use the word "intention" rather than "intent", as the former is not necessarily the output of a deliberative decision process. As a matter of fact, an intention can also result from a heuristic (intuitive, instinctive) decision process [31,32]. I can have a disposition to act that is due to an intuitive decision process, but this disposition might still be blocked, although maybe with more difficulty than dispositions to act that result from a deliberative decision process.

Intentions behave more like continuants than like occurrents: John's intention to read a book in May can be wholly present at different times in April, and wax and wane as time passes. Moreover, intentions have a dispositional character in the sense that my intention to do *A* typically leads to me doing *A*, or can be blocked. For example, I formed the intention to go for lunch but suddenly I remember that I have to write this email, so my intention is not realized. Or more radically, I decided to stand up but suddenly I'm paralyzed, so I cannot. Therefore, we categorize intentions as (mental) dispositions. In (HEATER), John's (dispositional) intention d_{INT} to activate the heater is realized in his action o_{ACT} of pushing the button of the heater (formally: d_{INT} **realized_in** o_{ACT}).

Of course, dispositions can exist without intentions: inanimate objects do have plenty of dispositions. For example, John also has a disposition **d** to push the button of the heater that is triggered by a heavy object pushing his finger down on the button, but this disposition is purely mechanical, and has nothing to do with his intention to push the button. The dispositions **d**$_{INT}$ and **d** do not have the same kinds of triggers: **d**$_{INT}$ is not triggered by a heavy object pushing down John's finger; hence those are two different dispositions, by Barton et al.'s [18] identity criterion for dispositions.

3.3.2. In favor of the non-reductivity of intentions to beliefs and/or desires

Our proposal presupposes that intentions are *bona fide* entities distinct from both beliefs and desires, in keeping with a vital theoretical role of intentions in commonsense psychology [6] and the "adequatist" principle of ontology building according to which "the entities in any given domain should be taken seriously on their own terms" [26, p. 46]. This non-reductive view of intentions has been nonetheless criticized. That is, some philosophical theories have identified intentions with beliefs, whereas others have identified intentions with desires, or with desire-belief compounds.

For instance, strong intention cognitivism maintains that intending to *V* consists in believing that one will *V*. It is primarily motivated by the linguistic observation that canonical sentences expressing intentions, such as "I am going to *V*" and "I will *V*" are also used to express beliefs [33]. As Levy [34] says, however, this linguistic argument is not convincing enough to establish an intimate connection between intentions and beliefs, since declarative sentences are typically used to express non-belief attitudes: e.g. "I'd like to know what time it is."

To take another example, a "belief-desire account" of intentions has been popular since the former Davidson's [35] theory of intentional actions. According to Mulder's [36] formulation of such account, an agent A intends to φ iff A desires, all things considered, φ. For example, Shihababu [37] argues that intentions are reducible to desires because desires can "motivate action when combined with an appropriate means-end belief". Alternatively, intentions could be identified with desire-belief compounds.

Indeed, there seems to be a close proximity between desires and intentions. Seen linguistically, for instance: "beliefs are like declarative sentences, which are satisfied (made true) by whether the world as it is conforms to them. But desires are like imperative sentences, which are satisfied (fulfilled) by changes in the world bringing the world into conformity with them" [29]. Intentions, like desires, are more like imperative sentences than declarative sentences: they are satisfied by changes in the world bringing it into conformity with them.

The crucial notion of practical reasoning nevertheless shows the distinction between desires and intentions [1,38]. However, as explained by the later Davidson [38], desires attach to actions less directly than intentions do. For example, John may have a desire to be warm, but this desire may be trumped by a stronger desire to spare energy. On the other hand, a (well-formed) intention to activate the heater sees as settled this issue of what action to perform so as to satisfy a desire to get warm. Mulder [36] argues that intentions need to be posited in order to capture the notion of practical reasoning that would evaluate and hierarchize the various beliefs and desires, and that issues in "a *practical judgment* about what is to be done". Our theory accounts for this non-reductive nature of intentions: occurrent desires and occurrent beliefs are parts of a decision process of practical reasoning, which may lead to the formation of an intention.

4. Formal ontological foundations of belief, desire, and intention

4.1. Core formalization

We will now propose some axioms in OWL. In the (ONLY) framework of disposition [17] presented in Section 2, we would say that a dispositional belief has as realization only occurrent beliefs, and that a dispositional desire has as realization only occurrent desires:

(Bel$_O$) *Dispositional belief* SubClassOf (**realized_in** only *Occurrent belief*)
(Des$_O$) *Dispositional desire* SubClassOf (**realized_in** only *Occurrent desire*)

Conversely, in (PARTHOOD), every maximal realization of a dispositional belief **bel** has as part some occurrent belief: intuitively, an occurrent belief is the "minimal part" that is to be found in every realization of a dispositional belief (but the realization might be larger, if, for example, the occurrent belief causes on its own other cognitive processes). That is:

(Bel$_P$) *Rmax*(**bel**) SubClassOf (**has_part** some *Occurrent belief*)

(note that we have to write one such axiom for every instance **bel** of *Dispositional belief*, which is a shortcoming for the OWL representation of the framework (PARTHOOD))

Similarly, we can state that every maximal realization of a dispositional desire **des** has as part some occurrent desire. That is:

(Des$_P$) *Rmax*(**des**) SubClassOf (**has_part** some *Occurrent desire*)

Let us now turn to decision processes and intentions. A decision process is a process that integrates some belief(s) and some desire(s) to yield an intention. We formalize it as stating that a decision process has as parts some occurrent belief(s) and occurrent desire(s):

(Dec) *Decision process* SubClassOf [(**has_part** some *Occurrent_belief*) and (**has_part** some *Occurrent_desire*)]

Every intention is the result of some decision process. This can be formalized using the **specified_output_of** relation from the Ontology for Biomedical Investigation (OBI) [39] as follows:

(Int-Dec) *Intention* SubClassOf (**specified_output_of** some *Decision_process*)

(note that the converse does not hold: a decision process may not lead to any intention, if, for example, the agent is still hesitant at the end of the decision process).

Moreover, an intention is realized in an action. In the (ONLY) framework, this would be written as:

(Int$_O$) *Intention* SubClassOf (**realized_in** only *Action*)

In the (PARTHOOD) framework, for every instance **i** of *Intention*, we would have:

(Int$_P$)   *Rmax*(**i**) SubClassOf (**has_part** some *Action*)

4.2. The dynamic structure of the decision process

In the former account, we did not enter into the details of how the decision process is structured. In particular, we did not specify how beliefs and desires could interact. One could imagine, for example, that an occurrent belief (e.g. that I'm allergic to apples) would trigger a dispositional desire (not to get an allergic crisis). In some cases at least, however, an occurrent desire could be triggered by something else than an occurrent belief (I can actively desire not to get an allergic crisis without having any active belief that I have an allergy).

Also, suppose for example that at t_1, John has both a desire to eat an apple and a desire to eat a peach. He deliberates whether he will eat an apple or a peach, and this decision process has as parts an occurrent desire to eat an apple, and an occurrent desire to eat a peach. How the weighing of desire takes place is a further question that exceeds this article.

4.3. Compatibility with additional axioms or hypotheses

We provided above the most basic conceptualization and formalization of the BDI entities. Our theory is compatible with some additional axioms or hypotheses, two of which we will discuss below: (i) some occurrent beliefs triggering some dispositional desires and (ii) intentions being always triggered by occurrent beliefs.

4.3.1. Occurrent belief triggering a dispositional desire

We did not delve into the details of the structure of a decision process, in particular exactly how beliefs and desires therein interact. We leave the determination of those interactions open, as they rely on complex psychological, neurological and epistemological questions that are out of scope of this article. Still, it is a plausible claim that one or more occurrent beliefs may trigger a dispositional desire. Suppose that John deliberates at t_0 whether he should press on the heater switch. His dispositional belief **d$_{BEL}$** will then be realized in his occurrent belief **o$_{BEL}$** that if he presses the switch, the temperature will increase. He therefore needs to consider whether he wants the temperature to increase. This will trigger his dispositional desire to be warm **d$_{DES}$**, that will be realized in an occurrent desire to be warm **o$_{DES}$**. In such a case, we would have a dispositional desire triggered by an occurrent belief: **d$_{DES}$ has_trigger o$_{BEL}$**.

4.3.2. Occurrent belief triggering an intention

To explain the other claim that an occurrent belief triggers an intention, let us consider Bosse et al.'s [40] following "semi-formal" (in their terminology) explanation of the BDI model in developing a recursive BDI-based agent model for the theory of mind:

- At any point in time: If a desire is present and a "belief in reason" is present, then an intention for an action will occur. (The term "reason" therein means "the (rational) choice of an action that is reasonable to fulfill the given desire".)
- At any point in time: If an intention for an action is present and a "belief in opportunity" is present, then the action will be performed. (The term "belief in

opportunity" therein refers to "the belief that certain circumstances in the world are fulfilled such that the opportunity to do the action is there".)

Not surprisingly, their first statement fits well with our view of a decision process as a process in which occurrent beliefs and occurrent desires closely interact in some way. Being possibly motivated by their recursive considerations, by contrast, their second claim can be formalized in our dispositional account of intention in such a way that an intention (disposition) would be stimulated by an occurrent belief (trigger), resulting in some action (realization). This is compatible with our general notion of triggers of dispositions and it can be formalized as follows:

Given (ONLY): *Intention* SubClassOf (**has_trigger** only *Occurrent belief*)

Given (PARTHOOD): For every instance **i** of *Intention*, *Tmin*(**i**) SubClassOf *Occurrent belief*

To illustrate this, suppose that a local singing contest takes place every month and Mary receives a regular voice training to win the competition while intending to participate in it when her voice will be trained to a certain level. When she believes at a certain time that she sings well enough to win the contest, she will perform the action of participating in the event of the month. In our ontological framework, Mary's occurrent belief in her developed singing skills triggers her intention to participate in a singing competition.

5. Discussion

5.1. Comparison with philosophical accounts of desires

It is well worth comparing our dispositional model of the BDI entities with some philosophical accounts of desires. Let us look at three major theories of desires in contemporary philosophy [29]. Note that we remain neutral on the object of desire that is designated by 'p' in the following (but see Section 5.5 for a brief discussion):

- **Action-based theory:** For an organism to desire p is for the organism to be disposed to take whatever actions it believes are likely to bring about p.
- **Pleasure-based theory:** For an organism to desire p is for the organism to be disposed to take pleasure in it seeming that p, and to take displeasure in it seeming that not-p.
- **Good-based theory:** For an organism to desire p is for the organism to believe that p is good.

Those theories embrace some kind of relationship between desires and dispositions. We will consider how each theory of desires can be formally characterized in OWL within our ontological framework for the BDI entities.

5.1.1. The action-based theory of desire

At first sight, the action-based theory of desire would dovetail with a deflationary view of intentions since it states that desires are something that disposes an agent towards actions, in contradistinction with our non-reductive approach to intentions. One possible

interpretation of this account reduces intention to desires, and considers that an intention is simply a desire which motivates an agent to act in a context where the agent has an appropriate "means-end belief": e.g. John's desire to warm up in the context where he has a belief that switching on a heater is the best way to achieve warmth [37]. The following axiom would ensue:

Intention SubClassOf *Dispositional desire*

In that case, we could keep $(Int_O)/(Int_P)$ but should reject $(Des_O)/(Des_P)$, since dispositional desires would be realized in actions, rather than in occurrent desires.

Another possible construal is that an intention is a desire-belief compound. Thus, an intention could be described as the mereological sum of a desire and a means-end belief:

Intention SubClassOf (**has_part** some *Dispositional desire*) and (**has_part** some *Dispositional belief*)

One question would be what kind of parthood (see [19] for three kinds of disposition-parthood) is involved in the axiom above. Answering this would be important to determine which of the former axioms from Section 4.1 would be accepted.

Let us now consider a more specific action-based theory. Ashwell [28,41] accounts for an agent's desire as a "second-order disposition" to have a particular behavioral disposition, such that which behavioral disposition is chosen is determined by her beliefs as to how she can bring about something desirable. Let us illustrate it with a variant of (HEATER). In Ashwell's framework, John's desire to get warm could be seen as a second-order disposition d^{2nd}_{DES} to acquire multiple behavioral dispositions, such as a disposition to activate a heater and a disposition to wear a sweater.

We will consider two ways to specify Ashwell's view, depending on how to interpret the terms "first-order" and "second-order" dispositions. The first one would be that a second-order disposition is a disposition that has two modes, and is therefore composed, in the "mod-parthood" sense of the term [19] mentioned earlier, by two first-order dispositions that each have a single causal pathway. Let $d_{activate}$ (resp. d_{wear}) be John's mental disposition realized in him activating the heater (resp. wearing his sweater) to get warm. We can think of the desire d^{2nd}_{DES} as a multi-track disposition [18,21] to activate the heater and to wear the sweater, whose realized pathway will depend on John's belief about which is the better way of warming him up. That is: both $d_{activate}$ **mod-part_of** d^{2nd}_{DES} and d_{wear} **mod-part_of** d^{2nd}_{DES} hold. Following Section 4.3.1, $d_{activate}$ (and thus d^{2nd}_{DES}) would be triggered by John's occurrent belief that activating the heater is the best way to get warm; whereas d_{wear} (and thus d^{2nd}_{DES} here too) would be triggered by his occurrent belief that wearing a sweater is the best way to get warm. In such a view, a desire would be realized by an action, and thus it should reject $(Des_O)/(Des_P)$.

According to the second interpretation of Ashwell's proposal, second-order dispositions are dispositions that are realized in a process leading to the formation of a first-order disposition. This would be in line with our own formalization presented in Section 4.1, which considers a desire as a disposition whose realization is part of a process that leads to the formation of an intention, which is a disposition to act. In this view of second-order dispositions, the intention is a first-order disposition to act, whereas desires (and beliefs) are second-order dispositions to act.

5.1.2. The pleasure-based theory of desire

Next, while leaving aside the psychological nature of (dis)pleasure, we can understand the pleasure-based theory as claiming that a dispositional desire (for *p*) is a disposition with two causal pathways: one to take pleasure in some circumstances (namely, seeming that *p*), and one to take displeasure in some other circumstances (namely, seeming that not-*p*). Such a disposition with two modes is formalized, as explained above, with the notion of "mod-parthood". Therefore, we could formalize the pleasure-based theory of desire based on two classes *Disposition to take pleasure* and *Disposition to take displeasure* and the **has_mod-part** relation (which we define as the inverse of **mod-part_of**), hence the following axiom:

> *Dispositional desire* SubClassOf (**has_mod-part** some *Disposition to take pleasure*) and (**has_mod-part** some *Disposition to take displeasure*)

In this framework, dispositional desires are realized by processes of taking pleasure or taking displeasure, rather than by occurrent desires that can be part of a decision process, as in our account; therefore, (Des$_O$)/(Des$_P$) should be rejected in such a framework.

5.1.3. The good-based theory of desire

A good-based theory of desire would typically consider that a dispositional desire for *p* is a dispositional belief that *p* is good, which leads to the following axiom:

> *Dispositional desire* SubClassOf *Dispositional belief*

Similarly, *Occurrent desire* might be seen as a subclass of *Occurrent belief*.

A slightly different construal would be that a desire of *p* would be a disposition to get a belief that *p* is good. Then, a dispositional desire for *p* is realized in an occurrent desire that has as output a dispositional belief that *p* is good. This can be formalized using the OBI:**has_specified_output** relation [39] and the (ONLY) model of dispositions as follows:

> *Dispositional desire* SubClassOf (**realized_in** only (*Occurrent desire* and **has_specified_output** some *Dispositional Belief*))

5.2. Belief-forming biases

Consider how our model could account for additional mental phenomena, such as cognitive biases. Let us take the example of wishful thinking [42], which we will understand as humans' tendency to believe what they desire. Suppose that John desires his lucky number to be selected at the lottery (manifested by his occurrent desire o^{lucky}_{DES}) and therefore believes that his lucky number will be selected at the lottery (hence his dispositional belief d^{lucky}_{BEL}). We would not have o^{lucky}_{DES} triggering d^{lucky}_{BEL}: indeed, d^{lucky}_{BEL} does not exist before o^{lucky}_{DES}, and thus cannot be triggered by it. Instead, we formalize the human bias of wishful thinking as another mental disposition d_{WISH} in Jonh's mind, a disposition to create wishful thinking beliefs on the basis of his beliefs. o^{lucky}_{DES} triggers d_{WISH}, which is then realized in a mental process of

wishful thinking (**o**WISH) that creates **d**^lucky^BEL (note that we do not claim that *all* occurrent beliefs would lead to the formation of a corresponding desire).

To represent formally **d**WISH, we might use the OBI relation **has_specified_output** as follows:

dWISH **has_trigger o**^lucky^DES
dWISH **realized_in** some (*Mental process* **and has_specified_output d**^lucky^BEL)

To generalize this, in the (ONLY) framework, we might introduce a new class *Dispositional_wishful_thinking* of cognitive biases as follows:

Dispositional_wishful_thinking SubClassOf (**has_trigger** only *Occurrent_desire*)
Dispositional_wishful_thinking SubClassOf [**realized_in** only (*Mental process* and **has_specified_output** some *Dispositional Belief*)]

5.3. Beliefs and moral motivation

Moral motivation is the motivation that is linked to one's moral (and thus normative) judgments. Rosati [43] describes it more systemically: When an agent judges that it would be morally right to perform an action, the agent is ordinarily motivated to perform the action. There is a philosophical debate over moral motivation between Humeanism and anti-Humeanism [43]. Humeanism says that the belief that acting in some way will lead to the achievement of a goal is not sufficient for motivation: a desire to achieve the goal is required as well. Anti-Humeanism counters that some beliefs, more specifically moral beliefs, can motivate an action on their own.

Let us identify here motivation with intention (although this would require further investigations). As said above, our model of the BDI entities states that intentions emerge from decision processes, which are characterized as processes in which not only beliefs but also desires interact in some way. Thus, anti-Humeanism would conflict with the combination of our axioms (Dec) and (Int-Dec), which imply together that all intentions to act result from a decision process that includes both (occurrent) beliefs and desires. Therefore, our framework is Humean. Note however that it would be easy to switch to a framework that would be agnostic concerning the issue of Humeanism by accepting, instead of (Dec), the following weaker axiom:

Decision process SubClassOf (**has_part** some *Occurrent_belief*)

5.4. Linking intentions with affordances

Gibson [44] defines affordances of the environment as "what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill" (p. 119). Canonical examples include the character of stairs to be potentially climbed and the character of gaps to potentially hide agents. A theory of affordances offers a first foundation upon which agents and inanimate objects (e.g. tools) are distinguished and identified [45]. According to Heras-Escribano's [46] affordance-based approach to agency, an agent's intentional actions are characterized by the agent's possibility to choose among different affordances.

Proposed by Turvey [47], a dispositional theory of affordances has been philosophically furthered by Heras-Escribano [46]. In formal ontology, it has been theoretically investigated by some of us [12-14], in a way that could also be used as a

basis for other accounts of affordances (e.g. [48]). For instance, the affordance of the stairs is their disposition to support people as they move upward (or downward) when the stairs are of a suitable proportion for their leg length. Those considerations could lead us to link intentions with affordances: for instance, Mary's intention to climb up the stairs is realized only when so is the affordance of the stairs to support her as she moves upward.

5.5. The content of the BDI entities and its aboutness

We make some brief observations about the content of the BDI entities that are to be taken seriously along with this line of inquiry. As we said in Section 3.1, the content of mental attitudes has been examined in philosophy under the name of "proposition" [22]. While the content of belief is traditionally acknowledged to be a proposition (as a "truth-bearer"), it is highly contentious whether desire is a "propositional attitude" as well [49]. When John desires to get warm, for instance, it is not obvious whether he desires a state of affairs [50] of him being warm (which is standardly taken to be propositionally structured) or he simply desires for warmth as a thing with no propositional structure. This discussion can be extended to the content of intention and also of many other mental attitudes such as love and fear [51].

It is important to remark that some existing ontologies of mental entities [3,7] share the idea that the notion of *aboutness* plays a vital role in connecting the BDI entities with their content. For instance, a previous work [11] on belief provides a preliminary formalization of the content of belief using the notion of information content entity [52] which is, by definition, about something. For another example, Biccheri et al. [53] complement a BDI approach to mental states (in their terminology) with their dual account of aboutness: a mental state is about an "intentional content" and also about an "intentional object". Aboutness still remains elusive in formal ontology, but it may be elucidated in terms of semiotics [54] which explicates meanings and representations in terms of the triad of a sign, an object and an interpreter.

6. Conclusion and future work

We examined beliefs, desires, and intentions conceptually and formally. Beliefs and desires are Janus-faced: dispositional beliefs (resp. desires) can be realized in occurrent beliefs (resp. desires). As distinct from beliefs and desires (or their compounds), intentions are dispositions to actions that emerge from a decision process in which occurrent beliefs and occurrent desires interact. This account can be linked to an action-based theory of desire such as Ashwell's [28,41], in which desires are interpreted as second-order dispositions to action, mediated by intentions (whereas pleasure-based or good-based accounts of desires would require some changes in the formalization). Our account can also mesh with some BDI model of agency such as Bosse et al.'s [40], and can represent some cognitive biases such as wishful thinking.

There are several ways in which we will be able to investigate agency, cognition, and actions. First, our discussion on desires needs to extend to undesirability: e.g. severe adverse effects from a new drug would be undesirable to me. This work will contribute to e.g. Grenier's [30] ontological analysis of risk as a disposition whose realizations would be undesirable for an agent. Second, it will require exploration exactly how our account of intentions can formalize e.g. what Searle [55] calls "intention in action": an intention that is not formed in advanced of the action and causes it by representing its

condition of satisfaction "on the fly". Third, the notion of agency is underpinned not only by affordances but also by image schemas: mental patterns that are extracted from the sensory and motile experiences [45]. A previous affordance-based ontological approach [14] to image schemas will help to consider the relationship between the BDI entities and image schemas. Such considerations are also linked to the notion of instrumental desire: e.g. Mary's desire to raise her feet because she has a desire to climb up the stairs. Fourth, cognitive and neuroscientific BDI-related studies can be integrated with our proposal e.g. by specifying the "categorical bases" [18] of dispositional belief and desire: for instance, John's dispositional belief has as its categorical basis some neural structure of his brain, just as fragility of this glass has as its categorical basis some molecular structure of the glass. Fifth and finally, it will be worthwhile to connect our ontology of the BDI entities with action-directing informational entities such as recipes [15]. This line of inquiry could yield e.g. further development of an ontology of drug prescriptions [56] in representing the processes of drug-taking that drug prescriptions can direct.

References

[1] Bratman ME. Intention, plans, and practical reason. Cambridge: Harvard University Press; 1987. 208 p.
[2] Wooldridge MJ. Reasoning about rational agents. Cambridge: MIT Press; 2000. 241 p.
[3] Ferrario R, Oltramari A. Towards a computational ontology of mind. In: Proceedings of FOIS2004. Amsterdam: IOS Press; c2004. p. 287-97.
[4] Trypuz R. Formal ontology of action: a unifying approach [dissertation]. University of Trento; 2007. 273 p. Available from: https://www.researchgate.net/publication/233990262
[5] Hobbs J, Gordon A. Goals in a formal theory of commonsense psychology. In: Proceedings of FOIS2010. Amsterdam: IOS Press; c2010. p. 59-72
[6] Kashima Y, McKintyre A, Clifford P. The category of the mind: folk psychology of belief, desire, and intention. Asian J. Soc. Psychol. 1998 Dec;1(3):289-313.
[7] Hastings J, Smith B, Ceusters W, Jensen M, Mulligan, K. Representing mental functioning: ontologies for mental health and disease. ICBO2012. c2012. p. 1-5.
[8] Schulz S, Jansen L. Towards an ontology of religious and spiritual belief. In: Proceedings of FOIS2018. Amsterdam: IOS Press; c2018. p. 235-60.
[9] Guizzardi G, Falbo RD, Guizzardi R. Grounding software domain ontologies in the Unified Foundational Ontology (UFO): The case of the ODE software process ontology. In: Proceedings of CIbSE2008. CIbSE; c2008. p. 127-40.
[10] Vacura M. Modeling artificial agents' actions in context - a deontic cognitive event ontology. Appl. Ontol. 2020; pre-pressed:1-35.
[11] Barton A, Duncan W, Toyoshima F, Ethier JF. First steps towards an ontology of belief. In: Proceedings of JOWO2018. CEUR Workshop Proceedings, vol. 2205; c2018. p. 1-5.
[12] Toyoshima F. Modeling affordances with dispositions. In: Proceedings of JOWO2018. CEUR Workshop Proceedings, vol. 2205; c2018. p. 1-6.
[13] Toyoshima F, Barton A. A formal representation of affordances as reciprocal dispositions. In: Proceedings of TriCoLore2018. CEUR Workshop Proceedings, vol. 2347; c2018. p. 1-14.
[14] Toyoshima F, Barton A. Linking image schemas with affordances: an ontological approach. In: Proceedings of JOWO2019. CEUR Workshop Proceedings, vol. 2518; c2019. p. 1-10.
[15] Barton A, Vieu L, Ethier JF. Directing actions. In: Proceedings of JOWO2019. CEUR Workshop Proceedings, vol. 2518; c2019. p. 1-10.
[16] Horridge M. Patel-Schneider PF. OWL 2 web ontology language Manchester syntax. 2nd ed. W3C Working Group Note 11 December 2012.
[17] Röhl J, Jansen. L. Representing dispositions. J. Biomed. Semant. 2011 Aug;2(Suppl 4):S4.
[18] Barton A, Grenier O, Jansen L, Ethier JF. The identity of dispositions. In: Proceedings of FOIS2018. Amsterdam: IOS Press; c2018. p. 113-25.
[19] Barton A, Jansen L, Ethier JF. A taxonomy of disposition-parthood. In: Proceedings of JOWO2017. CEUR Workshop Proceedings, vol. 2050; c2017. p. 1-10.
[20] Toyoshima F. Natural necessity: an introductory guide for ontologists. Appl. Ontol. 2020 Feb;15(1):61-89.

[21] Barton A, Jansen L. A modelling pattern for multi-track dispositions for life-science ontologies. In: Proceedings of ODLS2016. CEUR Workshop Proceedings, vol. 2692; c2016. p. H1-2.

[22] McGrath M, Frank D. Propositions. In: Zalta EN, editor. The Stanford encyclopedia of philosophy. Spring 2018 ed. Available from: https://plato.stanford.edu/archives/spr2018/entries/propositions/

[23] Schwitzgebel, E. Belief. In: Zalta EN, editor. The Stanford encyclopedia of philosophy. Fall 2019 ed. Available from: https://plato.stanford.edu/archives/fall2019/entries/belief/

[24] de Haas FAJ. Potentiality in Aristotle's psychology and ethics. In Engelhard K, Quante M, editors. Handbook of potentiality. The Netherlands: Springer; 2018. p. 71-91.

[25] Galton A, Mizoguchi R. The water falls but the waterfall does not fall: new perspectives on objects, processes and events. Appl. Ontol. 2009;4(2):71-107.

[26] Arp R, Smith B, Spear AD. Building ontologies with Basic Formal Ontology. MIT Press; 2015. 248 p.

[27] Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. In: Proceedings of the 2009 AMIA Summit on Translational Bioinformatics; c2009. p. 116-20.

[28] Ashwell L. The metaphysics of desire and dispositions. Philos. Compass. 2014 Jul;9(7):469-77.

[29] Schroeder T. Desire. In: Zalta EN, editor. The Stanford encyclopedia of philosophy. Summer 2017 ed. Available from: https://plato.stanford.edu/archives/sum2017/entries/desire/

[30] Grenier O. Ontologies du risque [master's thesis]. University of Sherbrooke; 2019. 150 p. Available from: http://hdl.handle.net/11143/15991

[31] Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica. 1979 Mar;47(2):263-91.

[32] Gigerenzer G, Gaissmaier W. Heuristic decision making. Annu. Rev. Psychol. 2011 Jan;62:451-82.

[33] Setiya K. Intention. In: Zalta EN, editor. The Stanford encyclopedia of philosophy. Fall 2018 ed. Available from: https://plato.stanford.edu/archives/fall2018/entries/intention/

[34] Levy Y. Why cognitivism? Can. J. Philos. 2018 Jun;48(2):223-44.

[35] Davidson D. Actions, reasons, and causes. J. Philos. 1963 Nov;60(23):685-700.

[36] Mulder JM. Why intentions? Ratio. 2018 Oct;31(S1):51-64.

[37] Sinhababu N. The desire-belief account of intention explains everything. Noûs. 2012 Jun;47(4):680-96.

[38] Davidson D. How is weakness of the will possible? In: Feinberg J, editor. Moral concepts. London: Oxford University Press; 1969. p. 93-113.

[39] Bandrowski A. et al. The ontology for biomedical investigations. PLoS One. 2016 Apr;11(4):e0154556.

[40] Bosse T, Memon ZA, Treur J. A recursive BDI-agent model for theory of mind and its applications. Appl. Artif. Intell. 2011 Jan;25(1):1-44.

[41] Ashwell L. Conflicts of desire: dispositions and the metaphysics of mind. In: Jacobs JD, editor. Causal Powers. Oxford: Oxford University Press; 2017. p. 167-76.

[42] Reber AS, Reber ES, Allen R. The penguin dictionary of psychology. 4th ed. Penguin Books; 2009. p. 928.

[43] Rosati CS. Moral motivation. In: Zalta EN, editor. The Stanford encyclopedia of philosophy. Winter 2016 ed. Available from: https://plato.stanford.edu/archives/win2016/entries/moral-motivation/

[44] Gibson JJ. The ecological approach to visual perception. Houghton Mifflin; 1979. 332 p.

[45] Kutz O, Troquard N, Hedblom MM, Porello D. The mouse and the ball: towards a cognitively-based and ontologically-grounded logic of agency. In: Proceedings of FOIS2018. Amsterdam: IOS Press; c2018. p. 141-8.

[46] Heras-Escribano M. The philosophy of affordances. Palgrave Macmillan; 2019. 232 p.

[47] Turvey MT. Affordances and prospective control: an outline of the ontology. Ecol. Psychol. 1992 Sep;4(3):173-87.

[48] Chemero A. An outline of a theory of affordances. Ecol. Psychol. 2003 Jan; 15(2):181-95.

[49] Thagard P. Desires are not propositional attitudes. Dialogue. 2006;45(1):151-56.

[50] Armstrong DM. A world of states of affairs. Cambridge: Cambridge University Press; 1997. 300 p.

[51] Grzankowski A, Montague M, editors. Non-propositional intentionality. Oxford: Oxford University Press; 2018. 320 p.

[52] Ceusters W, Smith B. Aboutness: towards foundations for the Information Artifact Ontology. In: Proceedings of ICBO2015. CEUR Workshop Proceedings, vol. 1515; c2015. p. 1-5.

[53] Biccheri L, Ferrario R, Porello D. Needs and intentionality: an ontological analysis and an application to public services. In: Proceedings of FOIS2020. Amsterdam: IOS Press; c2020. To appear.

[54] Peirce CS. The essential Peirce. vol. 2. Peirce edition project. Bloomington I.N.: Indiana University Press; 1998. 624 p.

[55] Searle J. Intentionality: an essay in the philosophy of mind. New York: Cambridge University Press; 1983. 278 p.

[56] Ethier JF, Barton A, Taseen R. An ontological analysis of drug prescriptions. Appl. Ontol. 2018 Nov;13(4):273-94.