Reasoning in versus about attitudes: forming versus discovering one's mental states

Franz Dietrich, Antonios Staras

Reasoning in versus about attitudes: Forming versus discovering one's mental states

This version: September 2021[1]

Franz Dietrich
Paris School of Economics & CNRS

Antonios Staras
University of Cardiff

Abstract

One reasons not just in beliefs, but also in intentions, preferences, and other attitudes. For instance, one forms preferences from preferences, or intentions from beliefs and preferences. Formal logic has proved useful for modelling reasoning in beliefs – a process of forming beliefs from beliefs. Can logic also model reasoning in multiple attitudes? We identify principled obstacles. Logic can model reasoning *about* one's attitudes – a process of discovering attitudes – but not reasoning *in* attitudes – a process of forming attitudes. Beliefs are special attitudes in that logic can model both reasoning about beliefs and reasoning in beliefs, namely as entailment between beliefs or entailment between belief *contents*, respectively. This makes beliefs the privileged target of logic as applied to psychology.

1 Introduction

A growing philosophical literature about rationality and practical reasoning teaches us that one reasons not only in beliefs, but in many other attitudes (e.g., Broome 2006, 2013, Kolodny 2005, 2007, Boghossian 2014). One can form preferences from preferences; such reasoning makes preferences more transitive. One can form the intention to help a child cross a street from the belief one ought to; such reasoning reduces akrasia. One can form the same intention from the intention to make the child happy and the belief the child's happiness requires the help; such reasoning increases instrumental rationality.

Attitudes can also change through other processes than reasoning, including processes driven by external causes (music can create desires) and internal psychological processes that are purely automatic and unconscious (desires can cancel intentions that stand in the way). But we focus exclusively on *reasoning*. We

[1]We are grateful for inspiring feedback from colleagues, notably from Robert Sugden and Frederik van de Putte. An earlier version had a different subtitle. Franz Dietrich acknowledges support from the French Research Agency through the grants ANR-17-CE26-0003, ANR-16-FRAL-0010 and ANR-17-EURE-0001.

adopt Broome's (2013) influential account of reasoning. There are other accounts, some of which count more mental processes as 'reasoning' than Broome's account.[2] Adjudicating between accounts is not our goal. We simply accept Broome's account.

Logic provides the predominant *formal* theory (or body of theories) of reasoning. Logic also provides powerful tools to model attitudes: modal operators, such as belief operators, preference operators, or intention operators. One would therefore conjecture that logic is able to model reasoning in multi-attitudes formally (modulo standard idealisations or abstractions that come with any formal model). That is, one would conjecture that reasoning in multi-attitudes (suitably idealised) follows an entailment relation of a suitable kind.

Truth of this conjecture is presupposed by the common conception of an ideal reasoner as someone who has reached a deductively closed set of attitudes. If reasoning need not follow entailment, then deductive closure need not be the characteristic mark of an ideal reasoner.

Our question is: to what extent is the conjecture or presupposition correct? Surprisingly, the conjecture turns out to be largely false. When logic engages with attitudes, it notoriously addresses something else, namely reasoning *about* attitudes. Examples are reasoning about preferences (e.g., Liu 2011), about beliefs (e.g., Halpern 2017), or about beliefs, desires and intentions (as in 'BDI logics'). Reasoning about attitudes lets one discover attitudes, not form them. It is a third-personal, meta-level process of attitude discovery, while reasoning in attitudes is an internal, first-personal process of attitude formation. In fact, reasoning about attitudes is a form of reasoning in beliefs: it is reasoning in *beliefs about attitudes*. It is theoretical reasoning whose objects are attitudes.

Reasoning in attitudes – which Broome calls reasoning 'with' attitudes[3] – differs fundamentally from reasoning about attitudes. Both matter in their own ways. Reasoning about attitudes matters where agents reason about one another in interactive settings (cf. Perea 2012), or reason about themselves in an act of reflection or introspection. It is a form of theoretical reasoning. Reasoning *in* attitudes is practical reasoning: it creates attitudes, including intentions that cause actions – a central part of mental activity. This makes such reasoning a natural subject for practical philosophy, psychology, and even artificial intelligence.[4]

[2] An example is Drucker's (forth.) broader account called 'generalism'.

[3] Our terminology might be better distinguishable from 'reasoning about attitudes'.

[4] Sophisticated intelligent systems use (artificial) reasoning to form (artificial) attitudes, including intentions that cause actions.

2 What is reasoning in attitudes?

To set the stage, this section discusses and formalises reasoning in attitudes, to the minimal extent needed here. The philosophical account follows Broome (2013), and the formalisation follows Dietrich et al. (2019).

2.1 Attitudes and constitutions

The agent – 'you' – holds various attitudes, also referred to as (mental) states, such as: believing it snows, desiring to feel warm, intending to dress warm, preferring snow to rain, etc. The set of all possible attitudes is denoted M. Those attitudes which you possess form your *(mental) constitution*, formally a subset $C \subseteq M$.

Think of attitudes in M as pairs of an attitude-content and an attitude-type. For many philosophers, contents are propositional (on propositionalism see Felappi forthcoming): they are *single* propositions for monadic attitudes like intention, *pairs* of propositions for dyadic attitudes like preference, etc. One could make this structure of states formally explicit.[5]

We use the term 'attitude' not only for mental states in M (such as: desiring to be warm), but also for attitude-types (such as: desire).

2.2 Reasoning, informally

Your constitution changes through reasoning. In reasoning, you form a (conclusion-)attitude from existing (premise-)attitudes: you form beliefs from beliefs; intentions from beliefs and desires; preferences from preferences; etc. The process is causal: the premise-attitudes cause the conclusion-attitude. It constitutes a conscious mental act. You bring the premise-attitudes to mind by saying their contents to yourself, normally using internal speech. This lets you construct a new attitude, again using (internal) speech. You might reason:

Paying taxes is legally required. So, I shall pay taxes. (1)

This is reasoning from a single premise (a belief) to an intention. The conclusion-attitude has this content: *I pay taxes.* What you say however involves 'shall', a linguistic marker indicating that you entertain the content as an intention. In reasoning, you express to yourself the *marked contents* of your premise- and conclusion-attitudes, not the contents simpliciter. Marked contents are contents marked by

[5] Let L be a set of *propositions*, and A a set of *attitude-types*, each carrying an *arity* $n \in \{1, 2, ...\}$, usually 1 (monatic attitudes) or 2 (dyadic attitudes). Plausibly, A contains at least belief *bel* (monadic), desire *des* (monadic), intention *int* (monadic), preference $\succ$ (dyadic), and indifference $\sim$ (dyadic). Finally, define *attitudes* in M as tuples $m = (p_1, ..., p_n, a)$ where a is an attitude type in A, n is its arity, and $p_1, ..., p_n$ are propositions in L. So, (p, bel) is believing p, (p, int) is intending p, $(p, q, \succ)$ is preferring p to q, etc.

how the content is entertained: as a belief, or intention, etc. The English language contains markers for various attitude types, allowing you to reason in those attitudes. Beliefs are special: they need no linguistic marker, as the same sentence – in the example: *Paying taxes is legally required* – expresses the belief's content and marked content.

Importantly, in reasoning you do not say to yourself *that you hold* the attitudes in question. You do not say:

I believe paying taxes is legally required. So, I intend to pay taxes.

This would be reasoning about your attitudes (cf. Section 3.2).

Reasoning is rule-governed: you draw the conclusion by following a *rule* that you endorse, although this endorsement is not an explicit act and requires no awareness of the rule, indeed of the concept of rules. A rule allows forming some (conclusion-)attitude from some existing (premise-)attitudes. Rules can be individuated differently. In its most specific individuation, the rule you follow in (1) is this: from believing that paying taxes is legally required, come to intend to pay taxes. In a broader individuation, the rule is a schema, such as: from believing that ϕ-ing is legally required, come to intend to ϕ (where ϕ is any act). Many rules promote your rationality. Here are examples of rationality-promoting rules, stated informally:

(a) *Modus-Ponens Rule*: From believing p and believing *if* p *then* q, come to believe q. Parameters: propositions p, q.

(b) *Enkratic Rule*: From believing *obligatorily* p, come to intend p. Parameter: propositions p.

(c) *Instrumental-Rationality Rule*: From intending p and believing q *is a means implied by* p, come to intend q. Parameters: propositions p, q.

(d) *Preference-Transitivity Rule*: from preferring p to q and preferring q to r, come to prefer p to r. Parameters: propositions p, q, r.

One could modify these rules, and add others. Exactly which rules you follow or should follow is not our topic.

It is debatable how exactly the English language expresses reasoning with these rules, i.e., which linguistic constructions serve to mark attitude-contents. Reasoning in preferences might at first seem obscure, as preferences are dyadic attitudes. Broome (2006) however points out (citing Jonathan Dancy for this insight) that English has a preference marker, namely a construction with 'rather'. You can reason in preferences as follows:

Rather bike than walk. Rather walk than drive. So, rather bike than drive.

You initially prefer biking to walking, and waking to driving. You come to prefer biking to driving using the Preference-Transitivity Rule, where p, q and r are *I bike*, *I walk* and *I drive*, respectively.

2.3 Reasoning, formally

As noted, rules can be individuated specifically or more broadly. Our official definition of 'rule' chooses the specific individuation. This choice simplifies the formalism; nothing hinges on it. So we define a **reasoning rule** as any specific combination (P, c) of a set of (premise-)attitudes $P \subseteq M$ and a (conclusion-)attitude $c \in M$. The four rule schemas (a)-(d) in Section 2.3 can now be re-stated:

- $(P, c) = (\{\textit{believing } p, \textit{believing } \textit{if } p \textit{ then } q\}, \textit{believing } q)$ for propositions p, q,
- $(P, c) = (\{\textit{believing obligatorily } p\}, \textit{intending } p)$ for propositions p,
- etc. for (c) and (d).

These re-statements are still semi-informal, but formal statements are possible.[6]

You reason with certain rules – 'your' rules. Henceforth, S denotes the set of your rules, your **reasoning system**. If you possess all premise-attitudes of a rule $r = (P, c)$ of yours, i.e., your constitution C includes P, then you can form the attitude c. Your new constitution is $C \cup \{c\}$. Should you already possess attitude c (i.e., $c \in C$), then your reasoning has merely 'reaffirmed' or 'refreshed' this attitude, and your constitution stays C $(= C \cup \{c\})$.

Starting from your initial constitution C, you can reason consecutively with your rules, thereby gradually forming new attitudes. This process converges to a constitution that is **stable under reasoning**, i.e., cannot change further through reasoning, as it contains the conclusion-attitude of each rule in S whose premise-attitudes it contains. This stable constitution – the endpoint of reasoning – does not depend on the order in which you apply your rules. It is denoted $C|S$ and called the **revision of C through reasoning**. Technically, $C|S$ is defined as the minimal extension of C stable under S.[7] Concretely, you reason into $C|S$ by first revising C through any rule $(P_1, c_1) \in S$ that is difference-making, i.e., has $P_1 \subseteq C$ and $c_1 \notin C$; then revising the result $C \cup \{c_1\}$ through another rule $(P_2, c_2) \in S$ that is difference-making, i.e., has $P_2 \subseteq C \cup \{c_1\}$ and $c_2 \notin C \cup \{c_1\}$; and so on until no difference-making rules remain. As long as S is finite, this process converges to $C|S$ after some (finite) number of steps. Our formal definition of $C|S$ also takes care of the case of infinite S.

[6] First use the formalism in footnote 5 to respectively write $(P, c) = (\{(p, bel), (\textit{if } p \textit{ then } q, bel)\}, (q, bel))$ $(p, q \in L)$, $(P, c) = (\{(\textit{obligatorily } p, bel)\}, (p, int))$ $(p \in L)$, etc. for (c) and (d). Finally, to give formal meaning to composite propositions, assume that to any propositions p, q is assigned a proposition $\textit{if } p \textit{ then } q$, to any proposition p is assigned a proposition *obligatorily* p, etc. Technically, this defines a binary operator $L \times L \to L$, a unary operator $L \to L$, etc. This makes the rules (a)-(d) formally well-defined. One could go further and model propositions in L syntactically (intensionally) as sentences in a formal language, or semantically (extensionally) as subsets of some set of possible worlds. This turns operators into syntactic or semantic operators, respectively (cf. Dietrich et al. 2020).

[7] Provably, this minimal extension exists, is unique, and equals the intersection of all stable extensions $C' \supseteq C$.

3 The difficulty to model reasoning in attitudes logically

It is tempting to try to model reasoning in attitudes through the (semantic or syntactic) entailment relation of a suitable formal logic. Surprisingly, there are principled obstacles. We now go (largely unsuccessfully) through the three most natural attempts to model reasoning in attitudes logically (Section 3.1–3.3). Later we return to the special case of reasoning in beliefs – theoretical reasoning (Section 4).

We shall assume throughout that reasoning is correct (error-free), to rule out trivial deviations of reasoning from entailment. So, your reasoning rules (in S) are correct. What exactly makes a rule correct is hard to say.[8] By a logical model of reasoning we mean throughout a model by an entailment relation. (This excludes another type of logical model provided by dynamic modal logics.[9])

3.1 Content entailment: a model of reasoning in a *single* attitude

The first attempt must be to model reasoning through entailments between attitude-*contents*. After all, this is what works to an important extent for reasoning in beliefs.

In the first place, logic is about propositions, not about anyone's attitudes. But propositions form the contents of attitudes. This opens the door for logicians to address attitudes. When logicians do so, they notoriously choose *beliefs*: they interpret propositions as *belief*-contents, which turns logical entailment into a model of reasoning in beliefs, not in desires, or in intentions, etc. One way to explain this 'belief bias' of logic is that beliefs share something with logic: beliefs are representational of an external reality, while desires or intentions are not. Beliefs should fit reality. Desires and intentions have the opposite direction of fit: reality should fit them. We return to beliefs in Section 4.

[8] Broome says little about it and appeals to intuition. Simple correctness criteria such as truth-preservation are unavailable outside reasoning in beliefs. The rules in (a)–(d) or versions thereof might be correct. The rule that goes from *desiring p* to *believing p* is incorrect, hence not in S. According to Broome, you reason correctly if you correctly follow correct rules, i.e., if (i) the rules you follow are correct and (ii) you make no mistake in following them. In our model, condition (i) means that S contains correct rules, and condition (ii) means that your constitution after reasoning is precisely $C|S$, which we assume.

[9] Such logics address attitude formation triggered by external events (e.g., public announcements), not internal reasoning. What is dynamic is not entailments, but (processes expressed by) certain sentences, e.g., 'after such-and-such, you believe such-and-such'. Yet we aim to model the reasoning process through entailments, not sentences. While standard modal logics model reasoning about attitudes, dynamic modal logics model reasoning about attitude *change*.

Could logicians instead choose desires (or intentions, etc.), and take entailments to model reasoning in desires (or intentions, etc.)? Such a model would support reasoning from desiring p into desiring p *or* q (or from intending p into intending p *or* q, etc.), as p entails p *or* q; and it would support reasoning from nothing into desiring a tautology (or intending it, etc.), as the empty set entails the tautology. One might doubt such reasoning, and hence reject that content entailment adequately models reasoning in desires (or in intentions, etc.). But even if content entailment successfully modelled reasoning in desires (or in intentions, etc.), we would not have modelled reasoning in *multi*-attitudes. Reasoning in desires (or in intentions, etc.) is still mono-attitude reasoning. Once we mix attitude types, as practical reasoning routinely does, content entailment obviously cannot model reasoning: while p and $if\ p\ then\ q$ entail q, you would not reason from desiring p and believing $if\ p\ then\ q$ into intending q.

In sum, the attempt to model reasoning through entailments between attitude-contents works for reasoning in beliefs (with qualifications discussed in Section 4), is debatable for reasoning in some fixed non-belief attitude such as desire or intention, and fails clearly for reasoning in multi-attitudes.

For even simpler reasons, content entailment cannot model reasoning in *non-monadic* attitudes, because such attitudes have complex contents. For instance, reasoning in preferences (Broome 2006) is reasoning in attitudes held towards *pairs* of propositions. Entailments go between propositions, not between pairs.

3.2 Attitude entailment: a model of reasoning *about* attitudes

We now turn to entailment between *attitudes*, rather than their contents. Read literally, attitude entailment models reasoning *about* attitudes, not *in* attitudes. Why? Assume you reason in attitudes as follows:

I ought to pay taxes. So, I shall pay taxes. (2)

Here you reason from a belief into an intention, following an instance of the Enkratic Rule in Section 2. One is tempted to model this reasoning by the entailment $B(p) \vdash I(q)$, where B is a belief operator, I is an intention operator, and p and q are sentences representing *I ought to pay taxes* and *I pay taxes*, respectively. Here and in other attitude entailments considered, we presuppose a suitable logic of attitudes, with modal operators for all attitude-types used in reasoning, such as a belief operator, an intention operator, or a (dyadic) preference operator. (Logics of attitudes exist in abundance. They can do many things.[10])

[10]Mono-modal logics address one attitude, e.g., belief in 'doxastic logics' (e.g., Halpern 2017) and preferences in 'preference logics' (e.g., Liu 2011). Multi-modal logics address more than one

Modelling your reasoning (2) by the entailment $B(p) \vdash I(q)$ is however problematic, because this entailment has a different literal reading:

I believe I ought to pay taxes. So, I intend to pay taxes. (3)

Here you reason *about* you attitudes: you deduce you have an intention, from having a belief. In (2) you do not reason about your attitudes; you might not even know you have them. You are not your own observer who notices a belief and deduces ('discovers') an intention. Instead you *form* an intention which did not exist. Attitude entailment models attitude discovery, not attitude formation. It models reasoning about attitudes, not in attitudes. Reasoning about attitudes does not change *these* attitudes, but it creates beliefs about them (cf. Broome 2014 and Dietrich et al. 2019).

Worse, the different reasoning (3) which the entailment models is invalid as an inference about your attitudes: its premise can hold without its conclusion. Indeed, before your true reasoning (2), you believed your ought to pay taxes without (yet) intending it; formally, $B(p)$ was true and $I(q)$ was false. Why, then, does the logic deem the inference $B(p) \vdash I(q)$ valid? Nothing is wrong with the logic, but we have misapplied it. The logic is designed for reasoning about *rational* attitudes, not *your* attitudes which can be akratic and still 'under construction'. Attitude entailment represents reasoning about *rational* attitudes, not about your attitudes while they are still irrational.

The point of reasoning *in* attitudes is to become more rational. Ironically, reasoning *about* your attitudes works if your attitudes are already rational, whereas reasoning *in* your attitudes matters if your attitudes are not yet rational. So, depending on whether your attitudes are already rational, you can reason about them or should reason in them.

To be precise, improving rationality need not be the purpose of reasoning in attitudes. You could for instance reason as in Section 2:

Paying taxes is legally required. So, I shall pay taxes. (4)

This is again reasoning from a belief into an intention. But it is not enkratic reasoning, because it starts from a belief about what is *legally* required, not what you ought to do. As Broome might say, you reason towards legality, not rationality. Modelling (4) by an entailment – namely by $B(p') \vdash I(q)$ where p' represents the new premise content – is again problematic, still because the entailment represents a piece of reasoning *about* your attitudes.[11] The novelty of the example is that we cannot use a logic of 'merely' rational attitudes. In such a logic, the entailment

attitude, e.g., belief, desire and intention in 'BDI logics'. Logics of attitudes capture rationality of attitudes by axioms (e.g., axioms requiring that tautologies are believed).

[11] The entailment reads: *I believe paying taxes is legally required. So, I intend to pay taxes.*

$B(p') \vdash I(q)$ would not exist, i.e., be invalid, because the premise-belief fails to *rationally* entail the conclusion-intention. But the logical invalidity of the entailment is an artifact of the choice of logic. The entailment $B(p') \vdash I(q)$ does hold under a logic of 'correct' attitudes in a suitably comprehensive sense of 'correctness' that captures all norms guiding your reasoning, such as rational, legal, or moral norms.

To sum up: attitude entailment in a suitable logic of rational or (more generally) 'correct' attitudes, models reasoning about rational or (more generally) 'correct' attitudes, but not reasoning *in* attitudes. Reasoning in attitudes matters if attitudes are not yet rational or (more generally) 'correct'.

3.3 Attitude entailment: an as-if model of reasoning in attitudes?

Could attitude entailment at least serve as an as-if model of reasoning in attitudes, instead of a literal model? That is, could attitude entailment correctly mimic the (external) attitudinal changes produced by reasoning in attitudes, whether or not it has anything to do with the (internal) psychological process at work? Reasoning in attitudes would then behave as if following attitude entailment, hence be *extensionally equivalent* to attitude entailment, in a sense made precise shortly. For instance, the entailment $B(p) \vdash I(q)$ in Section 3.2 would represent the reasoning in attitudes (2) *non-literally*, despite representing the reasoning about attitudes (3) *literally*. As-if interpretations of models are popular in rational-choice theory.[12] They remain the orthodoxy among economists, whilst being increasingly criticised (e.g., Dietrich and List 2016, Guala 2019).

We now formulate (and ultimately reject) the hypothesis of extensional equivalence between reasoning in attitudes and attitude entailment – the hypothesis underpinning a logical as-if model of reasoning in attitudes. Both sides of the (hypothetical) equivalence must be formalised:

- *The object to be modelled* is your (Broomean) reasoning in attitudes. Formally, it is captured by the transformation of constitutions $C \subseteq M$ into revised constitutions $C|S$, where S is your given reasoning system; $C|S$ contains the attitudes which are attainable by reasoning from your initial constitution C. For details see Section 2.
- *The object serving as as-if model* consists in entailments between attitudes. To formalise them, we presuppose a logic of attitudes in which attitudes are representable. Technically, each attitude $m \in M$ is represented by a sentence saying that you have this attitude; it is denoted m^* and takes the

[12] Under an as-if interpretation, a standard rational agent behaves *as if* maximising expected utility, so that utilities and probabilities carry no meaning beyond representing behaviour. Under more literal or mentalist interpretations, utilities and probabilities are psychological constructs capturing values and beliefs.

form $O(\phi)$ where O is the relevant attitude operator and ϕ is the relevant sentence. For instance, if m is *intending to swim*, then m^* is $O(\phi)$ where O is the intention operator and ϕ reads 'you swim'.[13] An entailment from your initial attitudes (in C) towards a new attitude c can now be formalised, as $\{m^* : m \in C\} \vdash c^*$.

We can now state the hypothesis:

Extensional Equivalence Hypothesis (EE): You can reason from your attitudes into an attitude if and only if your attitudes (represented logically) entail that attitude (represented logically). Formally, for all constitutions $C \subseteq M$ and all attitudes $c \in M$,

$$c \in C|S \Leftrightarrow \{m^* : m \in C\} \vdash c^*. \tag{5}$$

The left side of the equivalence (5) says that you can reason into c, starting from the constitution C. The right side says that the attitudes in C (represented logically) entail c (represented logically). In the example (3), c is *intending to pay taxes*, logically represented as $I(q)$ $(= c^*)$, and C contains *believing you ought to pay taxes*, logically represented as $B(p)$. Here EE says: you can reason into this intention if and only if your attitudes entail the intention; formally, $c \in C|S \Leftrightarrow \{m^* : m \in C\} \vdash I(q)$.

The current attempt to model reasoning *as if* following attitude entailment is non-literalist all the way. This is illustrated by some notable differences to the literalist approach tried (unsuccessfully) in Section 3.2. Section 3.2 focused on a specific instance of reasoning, namely (3). This instance is simplistic: you reason in just one step and you use just one premise. By contrast, EE is general: it captures the reachability of the attitude c by the condition $c \in C|S$, which is silent on how many reasoning steps (i.e., applications of a rule in S) are used to reach c, and which attitudes from C take part in the reasoning. The phenomenon modelled (the left side of the equivalence) is not an individual instance of reasoning, such as (3), but the possibility to reach an attitude from a constitution through reasoning, in any number of steps and using any of your attitudes as premises. The entailment that models the reasoning (the right side) starts from the totality of your attitudes, not just from some 'relevant' attitudes. In the example, the modelling entailment is not $B(p) \vdash I(q)$, but $\{m^* : m \in C\} \vdash I(q)$. The model does not reveal which of your initial attitudes produce the new attitude.

In sum, the as-if model based on EE treats reasoning as a black box that generates output attitudes from input constitutions, regardless of the psychological mechanism at work. This procedural blindness reflects the reduced ambition of the as-if approach, which aims to model what reasoning achieves *in effect*, not *how*

[13] Presumably, the assignment $m \mapsto m^*$ defines a bijective correspondence between M and the set of logical sentences of type $O(\phi)$ for some attitude operator O.

it achieves it – an approach we took reluctantly after the more substantive and mentalistic attempts had failed.

But is EE tenable, and with it the as-if model? We discuss an objection against sufficiency of entailment for reasoning (direction '$\Leftarrow$'), an objection against necessity (direction '$\Rightarrow$'), and a concern about ad-hoc-ness.

Against sufficiency. Sometimes you cannot form an attitude although your attitudes entail it. You might be akratic, and unable to form an intention c which rationally follows from your beliefs about what you ought to do. Here your constitution entails c (formally, $\{m^* : m \in C\} \vdash c^*$), but you cannot reason into c (formally, $c \notin C|S$). Or you believe that having attitude c makes happy; this belief (let us assume) rationally entails forming c, to which you are unable. Or your intend to become wise and believe studying is a necessary means, but you are psychologically unable to intend to study, although this intention is (rationally) entailed.

However, these counterexamples apply to an imperfect reasoner, who is unable to perform some correct reasoning. One can rehabilitate sufficiency by assuming a perfect reasoner who does not suffer from psychological 'reasoning barriers'. This reasoner's reasoning system S not only contains only correct rules (which we have assumed throughout), but also contains sufficiently many rules.

Might EE hold as a hypothesis about a perfect reasoner? Unfortunately not, since necessity fails.

Against necessity. You often reason into an attitude that does not follow from your attitudes. You might reason from intending to visit Venice and believing that Venice is reachable only by boat or train into intending to take a boat. Here your premise-attitudes do not entail your conclusion-intention, because rationality would have permitted to intend to take a train. (Recall: 'entails' means 'rationally entails', or more generally 'correctly entails' in a broader sense of correctness.) Two opposite reasonings are equally possible here: reasoning into intending to take the boat and reasoning into intending to take the train. Neither of these intentions, and certainly not both, are entailed by your current attitudes. As Broome (2013: 219) says, "[i]f it is correct to reason to some conclusion, that is because rationality permits you to reach that conclusion, not because it requires you to do so."

One might disagree, by mounting two claims: (i) you should not reason into a specific intention, but into the intention to take *either* a boat *or* a train; (ii) this disjunctive intention *is* rationally entailed by your premise-attitudes, i.e., by intending to visit Venice and believing that taking a boat or a train is a necessary means. But both claims seem problematic.

However, claim (i) begs the question of how you later form one of the two specific intentions (a specific intention being needed to reach Venice). The specific

intention is not entailed by the disjunctive intention or other existing attitudes. So, by EE, it does not emerge through reasoning; it emerges through another process, presumably some automatic causal process. But the idea that a disjunctive intention is formed through reasoning and is then automatically sharpened into a specific intention does not seem to match our real experience of decision-making. Intuitively, we reason directly towards a specific intention, without a detour over a disjunctive intention.

Claim (ii) seems incorrect. The two initial attitudes (of intending to visit Venice and believing that taking a boat or train is a necessary means) do not rationally require holding the disjunctive intention. They might rationally require holding some specific intention, i.e., intending to take the boat or intending to take the train; but the disjunctive intention is not required.

Ad-hoc-ness charge: Attitude entailments are entailments between attitude propositions of the simplest type: propositions saying that you possess some attitude, e.g., that you desire p. Call them *atomic* attitude propositions. There exist many *non-atomic* attitude propositions: that you do *not* desire p, that you desire p *and* believe q, etc. Entailments between non-atomic attitude propositions do not correspond to (Broomean) reasoning in attitudes. For instance, the entailment $\{B(p) \vee I(q), \neg D(r)\} \vdash \neg D(s)$ (for operators of belief B, intention I, and desire D) does not correspond to any reasoning in attitudes, because Broomean reasoning cannot start from disjunctions or absences of attitudes, and cannot result in absences of attitudes. You can reason *about* absences or disjunctions, but not *in* them (cf. Dietrich et al. 2019). It seems ad hoc to pick out particular entailments – those between *atomic* attitude propositions – and grant them a perfect correspondence to Broomean reasoning, while denying such a correspondence for all other entailments.

4 On the special status of reasoning in beliefs

Where do we stand? Reasoning in attitudes differs fundamentally from reasoning about attitudes. It lets you form rather than discover attitudes. It does not follow entailment. Neither does it follow entailment between attitude-*contents* – which models reasoning in beliefs (with some qualifications mentioned shortly). Nor does it follow entailment between *attitudes* – which models reasoning about your attitudes, by yourself or someone else.

The categorical difference between reasoning in and about attitudes holds even for beliefs only: reasoning in beliefs is not reasoning about beliefs.[14] But reasoning in beliefs (i.e., theoretical reasoning) stands out, because beliefs normally track an

[14]But the case against extensional equivalence (discussed later) might be weaker then

external truth: they normally aim to match the world, and are thus bound by logic. This is why theoretical reasoning follows entailment between attitude-*contents*.

Does it really? Theoretical reasoning sometimes departs from content entailment. You might derive *more* beliefs, by reasoning inductively. You might derive *fewer* beliefs, because subjectively probable (and believed) propositions sometimes jointly entail subjectively improbable (and disbelieved) propositions, as the Lottery Paradox illustrates. We say 'might' because our Broomean account of reasoning might escape at least the second objection, since explicit theoretical reasoning might exclude implicit probabilistic considerations.[15] Here is a third objection: arguably, theoretical reasoning can pursue non-epistemic goals, Arguably, reasoning in pursuit of non-epistemic goals can be correct reasoning. You can correctly reason into a belief in pursuit of happiness or because rationality requires having an opinion on some topic. If so, this would further disconnect theoretical reasoning from content entailment.

Still, content entailment is a first-order approximation of theoretical reasoning. Theoretical reasoning is thus approximately governed by logic, while reasoning in multi-attitudes goes beyond ordinary logic.

References

Boghossian, P. (2014) What is Reasoning? *Philos Stud* 169: 1–18

Broome, J. (2006) Reasoning with preferences? In: *Preferences and Well-Being*, S. Olsaretti ed., Cambridge University Press, 2006, pp. 183–208

Broome, J. (2013) *Rationality Through Reasoning*, Hoboken: Wiley

Dietrich, F., List, C. (2016) Mentalism versus Behaviourism in Economics: A Philosophy-of-science Perspective, *Economics & Philosophy* 32(2): 249-281

Dietrich, F., Staras, A., Sugden, R. (2019) A Broomean model of rationality and reasoning, *Journal of Philosophy* 116: 585-614

Dietrich, F., Staras, A., Sugden, R. (2020) Beyond belief: Logic in multiple attitudes, working paper

Drucker, D. (forthcoming) Reasoning beyond belief acquisition, *Noûs*, see https://doi.org/10.1111/nous.12363

Felappi, G. (forthcoming) Propositionalism and Questions that do not have Correct Answers, *Erkenntnis*, forthcoming. See https://doi.org/10.1007/s10670-021-00442-5

Guala, F. (2019) Preferences: neither behavioural nor mental, *Economics & Philosophy* 35(3): 383-401

Halpern, J. Y. (2017) *Reasoning About Uncertainty*, Cambridge, Massachusetts

[15] A Broomean reasoner can reason (explicitly) in *partial beliefs* (which of course will not follow entailment between the contents of partial beliefs). But this is not reasoning in (straight) beliefs.

& London: MIT Press

Kolodny, N. (2005) Why be rational? *Mind* 114: 509-563

Kolodny, N. (2007) State or process requirements? *Mind* 116: 371-385

Liu, Fenrong (2011) *Reasoning About Preference Dynamics*, Dordrecht: Springer

Perea, A. (2012) *Epistemic Game Theory: Reasoning and Choice*, Cambridge University Press

Staffel, J. (2013) Can there be reasoning with degrees of belief? *Synthese* 190(16): 3535-3551