Abstract rationality: the 'logical' structure of attitudes

March 2022

Franz Dietrich
Paris School of Econ. & CNRS

Antonios Staras
U. of Cardiff

Robert Sugden
U. of East Anglia

Abstract

We present an abstract model of rationality theories that focuses on structural properties of attitudes. We construe rationality as coherence between one's attitudes, e.g., one's beliefs, values, and intentions. We introduce three 'logical' conditions on attitudes: consistency, completeness, and closedness. They generalise the classic logical conditions on beliefs towards multiple attitudes, but contrast with standard rationality conditions such as transitivity for preferences, modus ponens for binary beliefs, additivity for probabilistic beliefs, and non-akrasia for intentions. We establish a formal correspondence between our three logical conditions and standard rationality conditions. Addressing John Broome's enquiry into the achievability of rationality through reasoning, we characterize the extent to which explicit reasoning can help one become more 'logical', i.e., acquire consistent, complete, or closed attitudes, respectively. Our analysis forms a bridge between rationality and logic, and enables logical talk about multi-attitude psychology.

1 Introduction

There exist various concrete theories and models of rationality. They differ, firstly, in the object that qualifies as (not) rational, which could be preferences, binary beliefs, probabilistic beliefs, intentions, strategies, or often combinations of these. They differ, secondly, in the rationality requirements on that object, which could for instance include transitivity for preferences, modus ponens for binary beliefs, or additivity for probabilistic beliefs. Some rationality requirements link attitudes of different types; for instance, 'EU rationality' requires preferences over uncertain prospects to respond rationally to probabilistic beliefs and outcome evaluations, game theoretic rationality requires strategies to respond rationally to preferences and beliefs about opponents, and according to many philosophers intentions should rationally respond to ought-beliefs to prevent akrasia.

In the face of such diversity, this paper aims for an abstract and unified model of rationality theories that focuses on the *structure* of rationality requirements rather

than their substantive nature. For instance, substantively different requirements such as transitivity for preferences and modus ponens for beliefs share the same structure: that of a 'closedness condition', which requires that holding certain attitudes implies holding a certain other attitude.

We model rationality as a property of the abstract set of one's 'attitudes', which could contain beliefs, desires, preferences, intentions, admirations, or indeed any kinds of attitudes. This matches Broome's (2007, 2013) philosophical notion of rationality, but is also compatible with standard rational-choice-theoretic models of rationality, which could indeed be recast within our formalism.[1] Broome's approach is prominent in contemporary philosophical theorising about rationality and reasoning (other approaches include Kolodny 2005 and Boghossian 2014).

We take inspiration from logic at two levels. Firstly, our move to abstraction within the theory of rationality is related in spirit to the move to abstraction within abstract logic in a Tarskian tradition. Just as we shall focus on the structure of rationality and abstract away the concrete nature of attitudes and requirements, so abstract logic focuses on the structure of logical constructs (such as consequence operators) and abstract away the concrete nature of the logic and its language.[2] Secondly, we shall introduce three conditions on the set of one's attitudes that have a logical flavour and will be called 'consistency', 'completeness' and 'closedness'. Their structure differs considerably from that of standard rationality requirements such as preference transitivity.

Our analysis proceeds in different steps. After setting the stage in Section 2, we introduce our three logical conditions on multi-attitude psychology, which we relate first to the concept of rationality in general (Section 3) and then to specific rationality requirements of standard types such as preference transitivity (Section 4). Each of the two relationships will culminate in a theorem. Section 5 then addresses a version of Broome's (2013) central question of whether one can reason towards more rational attitudes: can one reason towards more *logical* attitudes, i.e., towards consistent, or complete, or closed attitudes? A third theorem will provide a tentative answer. Finally, Section 6 compares our abstract model of multi-attitude psychology with concrete logics of attitudes, such as logics of preferences (e.g., Liu

[1] Standard rational-choice-theoretic models characterise the agent by mental constructs (such as preferences, beliefs, utilities, and strategies) and define rationality in terms of these mental constructs and their relations. The mental constructs characterising the agent can be remodelled as a set of attitudes in our sense. But our Broomean model departs from a different, radically behaviourist notion of rationality that can also be found in rational-choice theory and that defines rationality as a property of choice patterns (technically, a choice function) rather than of mental constructs, the role of which is downgraded to that of 'choice representations'. For the contrast between mentalist and behaviourist interpretations of rational-choice models, see Dietrich and List (2016).

[2] The move to abstraction in logic goes back at least to Alfred Tarski's work in the 1930s about consequence operators on an abstract set of sentences (e.g., Tarski 1956) and has since then evolved into many directions, including those of algebraic logic and algebraic semantics.

2011), of beliefs (e.g., Halpern 2017), or of beliefs, desires and intentions (e.g., Van der Hoek and Wooldridge 2003).

2 Attitudes and rationality

This section introduces basic concepts, following Broome's (2013) philosophical notion of rationality as formalised in Dietrich et al. (2019).

Attitudes. The agent – 'you' – holds several attitudes: beliefs, desires, preferences, intentions, etc. Let M be the non-empty set of all possible *attitudes*, also called *mental states*. M might contain: believing that it rains, believing that it is sunny, desiring to stay dry, intending to dress warmly, preferring sunshine to rain, etc. One might think of attitudes in M as pairs of an attitude content (an object) and an attitude type (such as belief, desire, or intention). Some attitudes in M could be graded (e.g., probabilistic beliefs[3] or graded desires[4]) or vague (e.g., vague probabilistic beliefs[5] or vague graded desires[6]). For most philosophers, contents are propositional: they are *single* propositions for monadic attitudes such as belief or desire, *pairs* of propositions for dyadic attitudes such as preference, etc. We shall say 'attitude' not only for mental states in M (such as: intention to swim), but occasionally also for attitude types (such as: intention).

Those attitudes in M which you possess form your *constitution*. Formally:

Definition 1 *A **(mental) constitution** is any set $C \subseteq M$ of mental states, 'your' states.*

The description of an agent in a choice-theoretic model can usually be recast as a constitution within our framework.[7]

Rationality. Certain constitutions count as 'rational', the others as 'irrational'.

[3] Believing with subjective probability 0.8 that it rains is an attitude with content 'it rains' and attitude type 'belief to the (probabilistic) degree 0.8'.

[4] Desiring some outcome to the degree 7 is an attitude with content this outcome and with attitude type 'desire of degree 7'. Some would call the degree of desire the 'utility'.

[5] A vague probabilistic belief in something is often captured by a non-empty probability interval $I \subseteq [0,1]$ (which becomes a sharp belief if I contains a single number). For instance, believing with vague subjective probability $[0.6, 0.8]$ that it rains is an attitude with content 'it rains' and attitude type 'belief of vague probabilistic degree $[0.6, 0.8]$'.

[6] A vague desire can be captured by a non-empty 'utility interval' $U \subseteq \mathbb{R}$. For instance, desiring an outcome to the vague degree $[10, 15]$ is an attitude with content the outcome and type 'desire of vague degree $[10, 15]$'.

[7] For instance, the description of a Savage agent by a probability measure, a utility function and a preference relation can be recast as a constitution consisting of probabilistic beliefs (of type believing event such-and-such to the probabilistic degree such-and-such), graded values/desires (of type valuing/desiring outcome such-and-such to the degree/utility such-and-such) and weak preferences (of type weakly preferring act such-and-such to act such-and-such).

We identify a notion or theory of rationality with the set of constitutions it deems rational. Formally:

Definition 2 *A **notion or theory of rationality** is a set $T \subseteq 2^M$ of ('rational') constitutions.*

Rationality notions in rational-choice theory can usually be recast within our framework.[8]

An illustration. In practice, theories of rationality can be defined by specifying conditions on attitudes. Rational constitutions are then constitutions satisfying these conditions. To give examples of conditions, let us first formalise the structure of states. Let L be a set of *propositions*. Let A be a set of *attitude types*, each one endowed with

- an *arity* $n \in \{1, 2, ...\}$, which is usually 1 (for unary or monadic attitudes) or 2 (for binary or dyadic attitudes), and
- a *domain* $D \subseteq L$ of possible objects (contents) of the attitude. For instance, the domain of intention is the set of propositions one can intend.[9]

A could contain the (monadic) attitudes types of belief *bel*, desire *des* and intention *int*, and the (dyadic) attitude types of preference $\succ$ and indifference $\sim$, each having certain domains. Let the states in M be the tuples $m = (p_1, ..., p_n, a)$ in which a is an attitude type in A, n is a's arity, and $p_1, ..., p_n$ are propositions from a's domain. For instance, (p, bel) is believing p, (p, int) is intending p, $(p, q, \succ)$ is preferring p to q, etc. Here are some typical conditions on your constitution C, more precisely condition schemas parameterized by propositions (where we assume for simplicity that each attitude type has full domain $D = L$, i.e., that any proposition can be believed, intended, etc.):

R1: *Modus Ponens*: Believing p and $if\ p\ then\ q$ implies believing q, formally $(p, bel), (if\ p\ then\ q, bel) \in C \Rightarrow (q, bel) \in C$. Parameters: $p, q \in L$.

R2: *Non-Contradictory Desires*: Desiring p excludes desiring *not* p, formally $(p, des) \in C \Rightarrow (not\ p, des) \notin C$. Parameter: $p \in L$.

R3: *Enkrasia (Non-Akrasia)*: Believing that *obligatorily* p implies intending p, formally $(obligatorily\ p, bel) \in C \Rightarrow (p, int) \in C$. Parameter: $p \in L$.

[8] For instance, the constitution of a Savage agent (see footnote 7) is rational in the expected-utility sense if and only if its beliefs obey the probability axioms and its preferences are linked to its beliefs and values through the expected-utility criterion. In a different formalisation of the Savage agent which suppresses beliefs and values, that agent's constitution consists only of weak preferences (not of beliefs or values) and is rational if and only if its preferences obey Savage's axioms.

[9] When recasting standard rational-choice models within the present framework, then an agent's different attitude types indeed have different domains. For instance, a Savage agent holds (probabilistic) beliefs about Savage events but preferences about Savage acts, and a player in a game holds beliefs about moves of other players but intentions (strategies) about own moves.

R4: *Instrumental Rationality*: intending p and believing q *is a means implied by* p implies intending q, formally $(p, int), (q \textit{ is a means implied by } p, bel) \in C \Rightarrow (q, int) \in C$. Parameters: $p, q \in L$.

R5: *Preference Transitivity*: preferring p to q and q to r implies preferring p to r, formally $(p, q, \succ), (q, r, \succ) \in C \Rightarrow (p, r, \succ) \in C$. Parameters: $p, q, r \in L$.

R6: *Preference Acyclicity*: you do not simultaneously prefer p_1 to p_2, p_2 to p_3, ..., p_{k-1} to p_k, and p_k to p_1, formally $(p_1, p_2, \succ), (p_2, p_3, \succ), ..., (p_{k-1}, p_k, \succ) \in C \Rightarrow (p_k, p_1, \succ) \notin C$. Parameters: any number $k \geq 1$ and any $p_1, ..., p_k \in L$.

R7: *Preference Completeness*: you have some preference or indifference between p and q, formally $(p, q, \succ) \in C$ or $(q, p, \succ) \in C$ or $(p, q, \sim) \in C$. Parameters: $p, q \in L$.

Are these conditions requirements of rationality? Should we refine their formulation? What else does rationality require? These important questions are not our topic. What matters for us is that any given list of conditions defines a theory of rationality: the theory T that deems (only) the constitutions satisfying these conditions rational. For instance, T could be defined as the set of constitutions $C \subseteq M$ satisfying R1–R7. This definition can of course only be plausible if the only attitude types in A are *bel*, *des*, *int*, $\succ$, and $\sim$. It is no longer plausible if, say, A also contains probabilistic beliefs, i.e., if for each $t \in [0, 1]$ A contains an attitude bel_t of belief with subjective probability t. For such probabilistic beliefs, rationality might require additivity and other properties.[10]

In stating R1–R7, we have implicitly assumed that certain composite propositions can be formed within L. Specifically, whenever L contains propositions p and q, L contains specific propositions *not* p, *if* p *then* q, *obligatorily* p, and q *is a means implied by* p.[11] Some readers might want to model propositions syntactically (intensionally), letting L contain the well-formed sentences of a suitable formal language. Then the mentioned composite propositions are composite sentences: *not* p stands for $\neg p$, *obligatorily* p stands for $O(p)$ where O is a sentential 'obligation' operator, etc. Other readers, especially economists, might want to model propositions semantically (extensionally), letting L contain subsets of a given set of possible worlds Ω. Here the mentioned composite propositions are constructed semantic-

[10] Additivity says: if you believe p to degree t and q to degree t' and p *or* q to degree t'' then $t'' = t + t'$, formally $(p, bel_t), (q, bel_{t'}), (p \textit{ or } q, bel_{t''}) \in C \Rightarrow t'' = t + t'$, with parameters any mutually inconsistent propositions $p, q \in L$ and any $t, t', t'' \in [0, 1]$. One might also require that no proposition is believed to two different degrees, and that tautologies can only be believed to degree 1. More demandingly, one might require existence of probabilistic beliefs about certain propositions (e.g., the 'events' in a Savage framework): you believe p to some degree, formally there exists a $t \in [0, 1]$ such that $(p, bel_t) \in C$, with parameter any proposition p from a given 'belief domain' $L_{bel} \subseteq L$. All these conditions are required under a standard Savagean expected-utility model of rationality.

[11] Technically, the assignments $p \mapsto$ *not* p and $p \mapsto$ *obligatorily* p define two unary operators $L \to L$, and the assignments $(p, q) \mapsto$ *if* p *then* q and $(p, q) \mapsto q$ *is a means implied by* p define two dyadic operators $L \times L \to L$.

ally: *not* p is the complement $\Omega \backslash p$, *obligatorily* p is $O(p)$, where O is a semantic 'obligation' operator mapping Ω-subsets to Ω-subsets, etc. If one wants to model propositions rather than letting them be primitive objects, then the choice between the syntactic and semantic models is to some extent a matter of taste and convenience; the semantic model is probably simpler, and certainly more coarse-grained.[12] The philosophical plausibility of each model depends on one's view about the nature of propositions.[13]

3 Three 'logical' conditions on attitudes

Can your attitudes commit a *logical* mistake? That is, are attitudes subject to requirements of a distinctively logical flavour, as opposed to common rationality requirements like those in R1–R7? We now introduce three logical conditions on attitudes. We call them 'consistency', 'completeness', and 'closedness' because they are multi-attitude counterparts of the equally-named logical conditions on beliefs.

The logical notions could be related to the rationality notion in two opposite ways: either rationality generates logical notions, or logical notions generate a rationality notion. We shall explore both approaches (Sections 3.1 and 3.2), and then compare them (Section 3.3), and finally discuss the special status of completeness (Sections 3.4). Appendix A relates our three logical notions to their standard belief-theoretic counterparts.

3.1 Top-down: from rationality to logical notions

The first way to model the three logical notions starts with a theory of rationality (given for instance by axioms such as R1–R7) and then constructs the logical notions. This can be done as follows:

[12] It cannot distinguish between logically equivalent propositions: *it neither snows nor rains* and *It is not the case that it snows or rains* correspond to the same set of worlds, hence to the same proposition. This can be problematic because attitudes often distinguish between equivalent propositions: we often believe or intend something without believing or intending something equivalent, for instance out of unawareness of the equivalence.

[13] The syntactic model of propositions is philosophically natural under a structural notion of proposition according to which propositions have an internal structure that parallels at least roughly that of sentences expressing them (although sentences may be more fine-grained). The semantic or set-theoretic model of propositions is philosophically natural under a non-structural notion of propositions. King (2019) reviews both notions of proposition. Whether a syntactic or semantic model of propositions is more plausible is also related to whether one has an intensional or extensional notion of proposition, i.e., whether one takes proposition to be intensions or extensions of sentences – but here we enter controversial questions about the nature of extension (reference, *Bedeutung*) and intension (meaning, *Sinn*). Under an arguably plausible view, a sentence's extension and intension is structurally similar to a set of worlds or the sentence itself, respectively. Under an arguably less plausible Fregean view, they are structurally like a truth value or a set of worlds, respectively; this other view makes the semantic model extensional, not intensional.

Definition 3 *Given a theory of rationality T, a constitution C is*

- ***consistent*** *if there exists a rational constitution $C' \supseteq C$,*
- ***complete*** *if there exists a rational constitution $C' \subseteq C$,*
- ***closed*** *if C contains each attitude in M that it (rationally) entails, where being* ***(rationally) entailed*** *or* ***T-entailed*** *by C means being contained in all rational constitutions $C' \supseteq C$.*

What is the intuition behind these definitions?

- *Consistency* means that your attitudes cohere with one another, i.e., do not rule out one another. You are permitted to have your attitudes simultaneously. You might be forbidden to hold *only* them; but you can hold *at least* them. For instance, suppose you intend p, believe q *is a means implied by* p, but do not intend q. Your constitution is then not rational, assuming rationality requires Enkrasia R5; but your constitution is consistent, as long as it could be made rational by adding suitable attitudes including the intention of q.
- *Completeness* means that you have 'enough' attitudes. Your attitudes do not require additional attitudes. You are permitted to have *no more* than your attitudes. You might be forbidden to hold *all* your attitudes; but you can hold *no more* than them. For instance, assume you prefer p to q and also prefer q to p. Then your constitution is not rational, assuming rationality requires Preference Acyclicity R6; but your constitution is complete, as long as it could be made rational by removing suitable attitudes, including one of the two mentioned preferences.
- *Closedness* means that you have each attitude that rationally follows from your attitudes. For instance, supposing rationality requires Instrumental Rationality R4, then believing *obligatorily* p rationally entails intending p; so, closed constitutions containing the mentioned belief also contain the intention. Closed constitutions need not be consistent or complete, let alone rational. For instance, the maximal constitution $C = M$ – where you believe everything, intend everything, etc. – is trivially closed, but it is irrational (in fact, inconsistent) under plausible theories of rationality. At the other extreme, the empty constitution $C = \varnothing$ – where you have no attitude whatsoever – is, under some theories of rationality, closed but irrational (in fact, incomplete).

3.2 Bottom-up: from logical notions to rationality

Under the previous approach, the three logical notions are children of rationality. We now take the opposite approach. We start from logical notions and derive a theory of rationality – like when logicians use logical notions to define which belief sets are rational. But what do we mean by logical notions, in all abstract generality?

Definition 4 *(a) A* ***consistency notion*** *is a set $CON \subseteq 2^M$ of ('consistent') constitutions such that whenever $C \in CON$ and $C' \subseteq C$ then $C' \in CON$ ('losing attitudes preserves consistency').*

(b) *A* ***completeness notion*** *is a set* $COM \subseteq 2^M$ *of ('complete') constitutions such that whenever* $C \in COM$ *and* $C \subseteq C'$ $(\subseteq M)$ *then* $C' \in COM$ *('gaining attitudes preserves completeness').*

(c) *A* ***closedness notion*** *is a set* $CLO \subseteq 2^M$ *of ('closed') constitutions that consists of all constitutions which are closed under some classical consequence operator, i.e., that equals* $\{C \subseteq M : C = Cn(C)\}$ *for some classical consequence operator* Cn *over* M.[14]

A consistency notion CON captures the absence of tensions between attitudes by *some* standard (and is accordingly closed under taking subsets). A completeness notion COM captures the presence of enough attitudes by *some* standard (and is accordingly closed under taking supersets). A closedness notion CLO captures the presence of all attitudes that follow from present attitudes by *some* standard (and is accordingly closed under *some* classical consequence operator).

In Section 3.1 we had defined special logical notions based on a theory of rationality T. We henceforth denote them by

$$\begin{aligned} CON_T &= \{C : C \subseteq C' \text{ for some } C' \in T\} \\ COM_T &= \{C : C \supseteq C' \text{ for some } C' \in T\} \\ CLO_T &= \{C : C \text{ contains all } m \in M \text{ s.t. } C \ T\text{-entails } m\}. \end{aligned}$$

These are indeed logical notions in the general sense of Definition 4, because CON_T is closed under taking subsets, COM_T is closed under taking supersets, and CLO_T consists of the closed constitutions under the consequence operator Cn_T that maps any $C \subseteq M$ to

$$\begin{aligned} Cn_T(C) &= \text{ set of attitudes } T\text{-entailed by } C \\ &= \{m \in M : m \text{ is in all } C' \in T \text{ s.t. } C \subseteq C'\} = \cap_{C' \in T: C \subseteq C'} C'. \end{aligned}$$

Deriving logical notions from a full-fledged theory of rationality T is a 'top-down' approach to logical notions. But under a 'bottom-up' approach, where could logical notions CON, COM and CLO come from? They could emerge from individual axioms about attitudes. For instance, the axiom schemas R1–R7 in our 'illustration' in Section 2 can serve to define logical notions, where we must carefully select the right axioms for each logical notion:

- A consistency notion CON can be defined by the consistency-type[15] schemas

[14] Recall that a *consequence operator* over (here) the set M is a function Cn mapping each set $C \subseteq M$ to a set $Cn(C) \subseteq M$ (of 'consequences' of C). It is called *classical* or a *closure operator* if it is inclusive ('$Cn(C) \supseteq C$'), monotonic ('$C \subseteq C' \Rightarrow Cn(C) \subseteq Cn(C')$'), and idempotent ('$Cn(Cn(C)) = Cn(C)$'). The classical consequence operator Cn underlying a given closedness notion CLO is unique, and maps each $C \subseteq M$ to its smallest extension in CLO. By this uniqueness, closedness notions and classical consequence operators are interdefinable.

[15] The expressions 'consistency-type schema','completeness-type schema', and 'closedness-type schema' should be intuitively clear. Technically, they denote schemas of, respectively, consistency conditions, completeness conditions, or closedness conditions, in a sense defined formally in Section 4.2.

R2 (Non-Contradictory Desires) and R6 (Preference Acyclicity). Formally, $CON = \{C : C \text{ satisfies R2 \& R6}\}$.

- A completeness notion COM can be defined by the completeness-type[16] schema R7 (Preference Completeness).
- A closedness notion CLO can be defined by the closedness-type[17] schemas R1 (Modus Ponens), R3 (Enkrasia), R4 (Instrumental Rationality), and R5 (Preference Transitivity).

Any logical notions can be used to define a theory of rationality:

Definition 5 *The theory of rationality* ***generated*** *by notions of consistency CON, completeness COM and closedness CLO is the theory that requires consistent, complete and closed constitutions, i.e., the theory $T = CON \cap COM \cap CLO$.*

One might wonder about the appropriateness of requiring completeness for rationality, given that preferences and logical beliefs are often not required to be complete. We discuss this issue in Section 3.4, but let us anticipate that this problem is only apparent since one can assume a vacuous completeness notion $COM = 2^M$, in which case rationality is effectively generated by consistency and closedness alone.

3.3 Comparing the top-down and bottom-up approaches to logical notions

We have considered two opposite approaches:

- Starting from a theory of rationality T and generating logical notions CON_T, COM_T, CLO_T.
- Starting from logical notions CON, COM, CLO and generating a theory of rationality $T = CON \cap COM \cap CLO$.

Interpretive differences aside, are both approaches formally equivalent? That is, are theories of rationality and logical notions interdefinable through some one-to-one correspondence? The answer is negative, for two reasons.

For one, some notions of rationality T are not reducible to any logical notions at all, because they are simply not structured along strict logical lines. So to say, rationality could go beyond logic. Rationality notions that *are* reducible to logical notions will be called 'classical', to highlight the parallel to classical notions of rational beliefs, which do indeed build on logical notions (cf. Appendix A). Formally:

Definition 6 *A theory of rationality T is* ***classical*** *if there exist notions of consistency CON, completeness COM and closedness CLO that generate T, i.e., satisfy $T = CON \cap COM \cap CLO$.*

[16] See footnote 15.

[17] See footnote 15.

For instance, the illustrative theory in Section 2, $T = \{C : C \text{ satisfies R1–R7}\}$, is classical, being generated by the consistency notion $CON = \{C : C \text{ satisfies R2 \& R6}\}$, the completeness notion $COM = \{C : C \text{ satisfies R7}\}$, and the closedness notion $CLO = \{C : C \text{ satisfies R1, R3, R4 \& R5}\}$ (cf. Section 3.2).

For another, one and the same (classical) notion of rationality can be generated by two different triples of logical notions.[18] So the 'true' logical notions are in general underdetermined by the notion of rationality. Yet, despite this underdetermination, one triple of logical notions stands out as canonical by consisting of the *logically strongest* logical notions that generate the given theory T. The canonical logical notions are precisely the logical notions CON_T, COM_T and CLO_T from the top-town approach in Section 3.1. We now formally state this result, proved in Appendix B.

Theorem 1 *For every classical theory of rationality T, CON_T, COM_T and CLO_T are the logically strongest consistency, completeness and closedness notions generating T, i.e., $T = CON_T \cap COM_T \cap CLO_T$ and all consistency, completeness and closedness notions CON, COM and CLO with $T = CON \cap COM \cap CLO$ satisfy $CON_T \subseteq CON$, $COM_T \subseteq COM$ and $CLO_T \subseteq CLO$.*

This result gives some salience to the logical notions from Section 3.1, and provides some support for the top-down approach to modelling logical notions.

3.4 Completeness – really?

In logic just as in rational-choice theory, completeness assumptions are often regarded as a matter of convenience rather than a requirement of rationality. Arguably, rationality does not require holding beliefs about everything, or preferences between any two options. Does this idea clash with our analysis that makes completeness a requirement of rationality? No, because our concept of completeness is very flexible. We allow the vacuous completeness notion 2^M, which deems all constitutions complete – even the empty constitution $C = \varnothing$. If one deems it permissible to hold no beliefs and no preferences, then one effectively endorses a completeness notion that does not require any beliefs or preferences. A more plausible completeness notion requires believing tautologies and being indifferent between options and themselves – but might not require anything else, as some would argue.

This highlights a departure of our abstract completeness notion from standard belief- or preference-theoretic completeness. 'Our' completeness is by definition

[18] For instance, for a fixed $m^* \in M$, the theory $T = \{C \subseteq M : m^* \notin C\}$ is the intersection of the consistency notion $CON = T$, the vacuous completeness notion $COM = 2^M$, and the vacuous closedness notion $CLO = 2^M$, but also the intersection of the notions of consistency $CON = 2^M \backslash \{M\}$, completeness $COM = 2^M$, and closedness $CLO = T \cup \{M\} = \{C : m^* \in C \Rightarrow C = M\}$.

rationally required but can be undemanding or even vacuous.[19] 'Standard' completeness (of beliefs or preferences) is very demanding but might not be rationally required. In principle, something similar applies to both other logical conditions: 'our' consistency and closedness are by definition rationally required but can be vacuous, whereas 'standard' consistency and closedness (of beliefs or preferences[20]) may or not be rationally required. But the contrast is smaller than for completeness, since the rationality of 'standard' consistency and completeness is less controversial.

Let us say this precisely, and add a terminological convention:

Remark 1 *Any of the logical notions CON, COM and CLO generating a (classical) theory of rationality T can be vacuous, i.e., equal to 2^M, in which case it drops out of the intersection $CON \cap COM \cap CLO$ defining T. Hereafter, vacuous logical notions can stay unmentioned when a theory of rationality is generated. For instance, we call a theory T* generated *by CON and CLO if $T = CON \cap CLO$, i.e., if T is generated by CON, 2^M and CLO.*

To honour the fact that standard theories of rationality often go without completeness requirement, let us call a classical theory 'fully classical' if it is generatable without completeness notion, i.e., with a vacuous completeness notion:

Definition 7 *A theory of rationality T is **fully classical** if there exist notions of consistency CON and closedness CLO that generate T, i.e., satisfy $T = CON \cap CLO$.*

Theorem 1 has a corollary for fully classical theories, as shown in Appendix B:

Corollary 1 *For every fully classical theory of rationality T, CON_T and CLO_T are the logically strongest consistency and closedness notions generating T, i.e., $T = CON_T \cap CLO_T$ and all consistency and closedness notions CON and CLO with $T = CON \cap CLO$ satisfy $CON_T \subseteq CON$ and $CLO_T \subseteq CLO$.*

4 Logical versus standard requirements of rationality

How are our logical conditions on multi-attitude psychology – consistency, completeness, closedness – related to standard conditions such as preference transitivity and the other conditions in R1–R7? To address this question, we must first settle on one of the two modelling approaches outlined in Section 3. Should our primitive object be a theory of rationality T or a triple of logical notions? Neither

[19] This is true under the top-down and bottom-up approaches to modelling completeness and the other logical notions (cf. Sections 3.1 and 4.2).

[20] By 'standard' consistency of preferences I mean acyclicity, and by 'standard' closedness of preferences I mean transitivity.

approach is fully general, because neither of the two objects generally determines the other. One might therefore reject both approaches and make both objects primitive, i.e., start with a primitive theory of rationality T *and* a primitive triple of logical notions CON, COM and CLO. One would then assume that the two objects are compatible, in the sense that rational constitutions satisfy the logical notions, i.e., that $T \subseteq CON \cap COM \cap CLO$. Assuming compatibility would be more general than assuming that the logical notions *generate* rationality, i.e., that $T = CON \cap COM \cap CLO$, or that rationality *generate* the logical notions, i.e., that $CON = CON_T$, $COM = COM_T$, and $CLO = CLO_T$. While interesting, this general approach will be set aside, for the sake of formal parsimony We shall also not let rationality be determined by logical notions, because this reductive approach would restrict us to *classical* theories of rationality – a limitation of generality we wish to avoid. Instead we shall make rationality our formal primitive, encouraged by the fact that, firstly, this approach leaves the theory of rationality T entirely general, and secondly, the logical notions CON_T, COM_T and CLO_T derived from the theory T, while not the only logical notions compatible with T, are somewhat canonical by Theorem 1.

So, the rest of the main text assumes that the notions of consistency, completeness and closedness are those determined by a given theory of rationality. The current section discusses the conceptual difference between logical and standard requirements of rationality (Section 4.1) and then presents a theorem that establishes a formal correspondence between both types of requirement (Section 4.2).

4.1 The conceptual difference between logical and standard requirements

Our logical requirements and standard rationality requirements like those in R1–R7 share an obvious feature: both are rationality requirements. Let us spell this fact out formally.

Definition 8 *A **condition** is a constraint on constitutions, formally a set $R \subseteq 2^M$ of constitutions (those 'satisfying' the condition). A condition R is a **requirement** of a theory of rationality T – for short, a **rationality requirement** – if it is satisfied by all rational constitutions, i.e., if $T \subseteq R$.*

Remark 2 *The three logical conditions CON_T, COM_T and CLO_T given by a theory of rationality T are rationality requirements.*

Having made this trivial point, let us see how logical and standard requirements *differ*.

1. Abstract versus concrete. Logical requirements are abstract and structural, since their definitions do not refer to the type or content of attitudes, but to structural relations between attitudes. Standard rationality requirements are concrete

and attitude-specific, since they are defined in terms of particular attitudes, such as preferences (in R5–R7) or intentions and beliefs (in R3–R4).

2. Global versus local. Logical requirements are global: they affect the constitution as a whole. Standard requirements are local: they concern only the (non-)possession of certain attitudes, regardless of the rest of the constitution. They are effectively constraints on a small subset of the constitution C. For instance, an instance of Preference Transitivity R5 concerns only C's intersection with $\{(p, q, \succ), (q, r, \succ), (p, r, \succ)\}$, and an instance of Enkrasia R3 concerns only C's intersection with $\{(obligatorily\ p, bel), (p, int)\}$. Christensen (2004) draws a similar global/local distinction, but for beliefs only.

3. Rationality-determined versus rationality-determining. This difference arises only under our current top-down approach of modelling logical notions as derivative objects; it should therefore not be universalised. While logical requirements are (under the top-down approach) determined by rationality, standard rationality requirements typically determine rationality. For instance, the 'illustration' in Section 2 invokes schemas R1–R7 of standard requirements that jointly determine or define a theory of rationality, which in turn determines or defines logical requirements. This striking difference in status or priority between standard and logical requirements could be given thinner or heavier meanings depending on what is read into 'determining'. Possible interpretations range from a mere functional or supervenience relationship between both objects to an explanatory relationship or even a relationship of metaphysical grounding.

4.2 The formal correspondence between logical and standard requirements

Despite all differences, logical and standard requirements of rationality stand in a tight formal relationship: each logical requirement is equivalent to a particular class of rationality requirements of standard type. But first, what are rationality requirements of standard type? A simple inspection of the rationality requirements discussed in philosophy or choice theory reveals that most of them, including those in the schemas R1–R7, fall into a three-kind typology. This typology is implicit in the work of Broome and others and formally introduced in Dietrich et al. (2019):

Definition 9 *The three **standard** types of condition consist of the following conditions, respectively:*

*(1) A **consistency condition** R forbids having all of certain attitudes, i.e., $R = \{C : not\ F \subseteq C\}$ for some set $F \neq \varnothing$ of attitudes (the 'forbidden set').*

*(2) A **completeness condition** R forbids having none of certain attitudes, i.e., $R = \{C : not\ C \cap U = \varnothing\}$ for some set $U \neq \varnothing$ of attitudes (the 'unavoidable set').*

*(3) A **closedness condition** R demands that having certain attitudes implies having a certain attitude, i.e., $R = \{C : P \subseteq C \Rightarrow c \in C\}$ for some set of ('premise-')attitudes P and some ('conclusion-')attitude c.*

The conditions in R1–R7 fall into this typology:

- Non-Contradictory Desires R2 and Preference Acyclicity R6 are schemas of consistency conditions, with forbidden set $\{(p, des), (not\ p, des)\}$ or $\{(p_1, p_2, \succ), (p_2, p_3, \succ), ..., (p_{k-1}, p_k, \succ), (p_k, p_1, \succ)\}$, respectively.
- Preference Completeness R7 is a schema of completeness conditions, with unavoidable set $\{(p, q, \succ), (q, p, \succ), (p, q, \sim)\}$. Another schema of completeness conditions is the schema in footnote 15, an instance of which requires holding at least some probabilistic belief in a given proposition.
- Modus Ponens R1, Enkrasia R3, Instrumental Rationality R4, and Preference Transitivity R5 are schemas of closedness conditions. In R1, the set of premise-attitudes is $\{(p, bel), (if\ p\ then\ q, bel)\}$ and the conclusion-attitude is (q, bel).

Having formalised logical conditions as well as conditions of standard type, we are ready to state the formal relationship between both kinds of condition. A tight correspondence holds by the following theorem.

Definition 10 *A **consistency/completeness/closedness requirement** of a theory of rationality T is a consistency/completeness/closedness condition that is a requirement of T (i.e., satisfies $T \subseteq R$).*

Theorem 2 *Given any theory of rationality $T \neq \varnothing$, a constitution C is*

(a) logically consistent if and only if it satisfies all consistency requirements of T,
(b) logically complete if and only if it satisfies all completeness requirements of T,
(c) logically closed if and only if it satisfies all closedness requirements of T,
(d) fully rational if and only if it satisfies all requirements of T.

Parts (a)–(c) connect the logical world of abstract requirements to the choice-theoretic or philosophical world of rationality requirements of standard type. Part (d) is an addendum, of interest in its own right.

Figure 1 displays schematically the requirements of a typical theory of rationality T. As usual in choice theory, the theory has been constructed from some set $\mathcal{A}$ of basic principles or 'axioms', for example the instances of the schemas R1–R7. That is, the rational constitutions are the constitutions satisfying the conditions in $\mathcal{A}$:

$$T = T(\mathcal{A}) = \{C : C \in R \text{ for all } R \in \mathcal{A}\} = \cap_{R \in \mathcal{A}} R.$$

Let all axioms be of a standard type: $\mathcal{A}$ consists of consistency conditions, completeness conditions, and closedness conditions. Of course, some other theories have *no* axiom of one of the three standard types (e.g., no completeness axiom) or have *additional* axioms of non-standard type; but in Figure 1 all three standard types, and only these types, occur among the axioms. The theory implies plenty of other

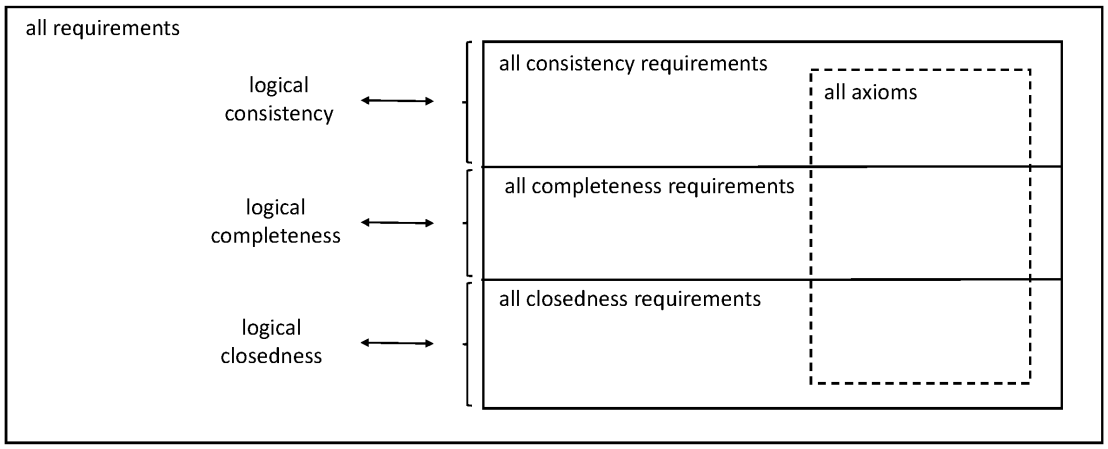

Figure 1: The rationality requirements of the theory $T(\mathcal{A})$

requirements besides the axioms. As indicated by the different areas in Figure 1, some additional requirements still fall within the standard typology.[21] The most salient requirements *outside* the typology are perhaps the three logical requirements; each of them is equivalent to a class of requirements of standard type by Theorem 2, as the arrows '$\leftrightarrow$' in Figure 1 indicate. Other requirements outside the typology are often artificial, including conjunctions or disjunctions of axioms.

Since the set of axioms $\mathcal{A}$ can be partitioned into sets $\mathcal{A}_{\text{con}}$, $\mathcal{A}_{\text{com}}$ and $\mathcal{A}_{\text{clo}}$ of consistency, completeness, or closedness conditions, respectively, the resulting theory of rationality $T = T(\mathcal{A})$ is classical, being generated by (i.e., the intersection of) the logical notions

$$CON = \{C : C \in R \text{ for all } R \in \mathcal{A}_{\text{con}}\}$$
$$COM = \{C : C \in R \text{ for all } R \in \mathcal{A}_{\text{com}}\}$$
$$CLO = \{C : C \in R \text{ for all } R \in \mathcal{A}_{\text{clo}}\}.$$

The theory would even be fully classical if, unlike in Figure 1, we had an empty set of closedness axioms $\mathcal{A}_{\text{com}} = \varnothing$ and hence a vacuous completeness notion $COM = 2^M$.

5 Reasoning towards logical requirements

So far our analysis was purely static, by focusing on coherence at a given time. But that static analysis has implications for the dynamic process of reasoning, as will be seen. We shall adopt Broome's account of reasoning, in the formal rendition of Dietrich et al. (2019).

[21] For instance, if $\mathcal{A}$ includes Preference Transitivity R5, then the theory $T(\mathcal{A})$ implies the following schema of closedness requirement (similar to R5 but with four propositions)

R5* preferring p to q and q to r and r to s implies preferring p to s, formally, $(p, q, \succ), (q, r, \succ), (r, s, \succ) \in C \Rightarrow (p, s, \succ) \in C$. Parameters: $p, q, r, s \in L$.

Instances of R5* are indeed requirements because, whenever $(p, q, \succ), (q, r, \succ), (r, s, \succ) \in C$, then $(p, r, \succ) \in C$ by R5 applied to p, q, r, and thus $(p, s, \succ) \in C$ by R5 applied to p, r, s.

In real life, your constitution is usually inconsistent, incomplete, and unclosed. Can reasoning help achieve these logical requirements? This question is a cousin of Broome's central question: can reasoning help achieve standard rationality requirements, such as instances of Preference Transitivity or Enkrasia? The tight connection between standard and logical requirements (Theorem 2) suggests an equally tight connection between Broome's and our question, i.e., between whether reasoning makes 'more (standardly) rational' and whether it makes 'more logical'. We shall confirm this conjecture.

Unlike Broome, we set aside whether reasoning is *correct* in some objective sense. Our conclusions about becoming 'more logical' through (Broomean) reasoning will be largely negative, and would get further reinforced by excluding incorrect reasoning. Our conclusions – including their negativity – depend strongly on our Broomean notion of reasoning, which we shall first introduce (Section 5.1) and compare with broader notions of reasoning (Section 5.2), before presenting our theorem (Section 5.3).

5.1 Reasoning in attitudes

Let us first introduce the Broomean notion of reasoning and formalise it following Dietrich et al. (2019). For Broome (2013), reasoning is a process of forming attitudes from existing attitudes: forming beliefs from beliefs, or intentions from beliefs and intentions, or preferences from preferences, etc. The process is causal. Unlike other causal processes, it is conscious and constitutes a mental act. It is explicit: you bring the premise-attitudes to mind by 'saying' their contents to yourself, usually through internal speech, which causes you to 'construct' and thereby acquire some conclusion-attitude, again using (usually internal) speech.

Here is a stylised instance of reasoning with a single premise. You say this to yourself:

Doctors recommend resting. So, I shall rest.

This is reasoning from a belief into an intention. The 'So' is not part of the conclusion, but expresses the act of drawing the conclusion. In reasoning, you say to yourself, not contents of attitudes simpliciter, but marked contents, i.e., contents with a marker indicating *how* you entertain the content: as a belief, or an intention, etc. In reaching the intention with content 'I rest', you say 'I *shall* rest', using '*shall*' as a marker for intention. The English language provides markers for various attitudes, including desire and preference (Broome 2013). Beliefs are special: they need no explicit marker (in English), as the same sentence expresses the content and the marked content.

Reasoning is rule-governed: you draw the conclusion by following a *rule*. Rules can be individuated more or less broadly. In the example, the rule could be

- specific: from believing that doctors recommend resting, towards intending to rest.

- broader: from believing that doctors recommend ϕ-ing towards intending to ϕ. Parameter: any act ϕ.
- even broader: from believing that expert E recommends ϕ-ing towards intending to ϕ. Parameters: any expert E and act ϕ.

We will work with specific rules, to avoid dealing with schemas and parameters. Nothing hinges on this technical choice: our results could be re-stated (more clumsily) using a broader notion of rule. Given our choice, we identify a rule with a specific premises/conclusion combination. Technically, a **reasoning rule** is a pair (P, c) of a set of ('premise-')attitudes $P \subseteq M$ and a ('conclusion-')attitude $c \in M$, representing the formation of c from P. In the rule in the example above, P contains just believing that doctors recommend resting, and c is intending to rest.

You can follow certain rules – 'your' rules. Your rules are those rules that capture premise-to-conclusion processes that you endorse, i.e., that (in Broome's words) seem right to you. The totality of your rules is your 'reasoning system', representing your reasoning policy. Technically, a **reasoning system** is a set S of reasoning rules. Starting from your initial constitution, you can reason with your rules: whenever you have a rule's premise-attitudes, you can form the rule's conclusion-attitude, which is added to your constitution. You can do this until your constitution is stable. A constitution C is **stable under** S ('**under reasoning**') if reasoning makes no change, i.e., C already contains the conclusion-attitude of each rule in S whose premise-attitudes it contains. The stable constitution reached by reasoning from your initial constitution C using your reasoning system S is denoted $C|S$ and called the **revision of** C **through** S ('**through reasoning**'). Technically, $C|S$ is defined as the minimal extension of C stable under S.[22] Provided your reasoning system S is finite, you can reach $C|S$ is in finitely many reasoning steps. You first apply a rule (P, c) in S that is effective ('difference-making') on C, i.e., for which $P \subseteq C$ but $c \notin C$; your constitution becomes $C \cup \{c\}$. You then apply another rule (P', c') in S that is effective on $C \cup \{c\}$; your constitution becomes $C \cup \{c, c'\}$. You continue until all your rules are ineffective. The order in which you reason, i.e., apply rules, is irrelevant: you inevitably converge to the same stable constitution $C|S$. All this can be stated formally.[23] For infinite reasoning systems S, one might object against our definition of $C|S$ that $C|S$ is defined too largely, as including even attitudes that are reachable only in 'infinitely many steps' (so to say). Admittedly, for infinite S our definition of $C|S$ is an idealisation, suitable for

[22] This (with respect to set-inclusion) minimal stable extension exists and is unique. It is the intersection of all stable extensions $C' \supseteq C$.

[23] Write $C|r_1|r_2|\cdots|r_n$ for the result of revising C through rule r_1, then through rule r_2, etc. until r_n. For finite S, $C|S$ can be shown to equal $C|r_1|r_2|\cdots|r_n$ for any sequence $(r_1, ..., r_n)$ of S-rules that is maximal subject to each rule r_i being effective on the previously reached constitution $C|r_1|r_2|\cdots|r_{i-1}$. In this representation of $C|S$ through consecutive reasoning, the sequence $(r_1, ..., r_n)$ (the way to reason) is only to a limited extent unique: all such sequences $(r_1, ..., r_n)$ have the same length (number of reasoning steps) n and the same set of conclusion-attitudes $\{c :$ some of $r_1, ..., r_n$ concludes in $c\}$.

infinite reasoners but not real reasoners.

5.2 Comparison with broader accounts of reasoning

Our Broomean account of reasoning differs from other accounts. We now discuss some key differences; our formal result will hinge on them.

Broomean reasoning is broad in that it operates within general attitudes, not just beliefs. But it is narrow in that it (i) forms but never removes attitudes, and (ii) is based on the presence but never the absence of attitudes. In short, you cannot reason to, or from, absences; you for instance cannot reason from *not* believing something to *no longer* intending something. These two features make the Broomean reasoning operator *inclusive* and *monotonic*.[24]

But our Broomean approach does not deny mental processes that produce, or start from, absences of attitudes, such as processes of losing intentions based on lacking certain beliefs. On the contrary, our Broomean approach regards such processes as a central element of psychology: an automatic element distinct from reasoning. Such automatic processes help improve rationality where reasoning alone is unsuccessful (Broome 2013 and Dietrich et al. 2019). This idea will be confirmed in Section 5.3.

But let us mention possible criticisms of our Broomean account of reasoning. For one, as this account precludes reasoning to or from absences, it seems to clash with belief elimination in AGM-type belief revision theory (Alchourrón et al. 1985) and with non-monotonic logics (Horty 2001). One might try to reconcile Broomean reasoning with these formal developments by interpreting AGM-type belief revision and non-monotonic logical consequence as capturing not reasoning alone but a combination of reasoning and automatic mental processes. We cannot explore here whether and how such a reconciliation works.

At a more philosophical level, Drucker (2021) has recently challenged Broome's concept of reasoning, defending a broader concept (based on Boghossian 2018). He argues that reasoning can not just add, but also remove attitudes. Roughly, according to his central thesis called 'Argumentalism', you reason towards an arbitrary attitudinal change (e.g., an attitude loss) when you run an argument that convinces you and that ends with a conclusion whose utterance expresses this change. For instance, suppose you have the initial belief that it rains. You reason towards losing that belief if you run an argument that convinces you and that concludes that it

[24] In our framework, a *reasoning operator* can be defined as any function transforming each initial constitution $C \subseteq M$ into a post-reasoning constitution $C^* \subseteq M$. In particular, our Broomean reasoning operator transforms each C into $C^* = C|S$, the revision of C under your (fixed) reasoning system S. This special reasoning operator obeys two axioms. *Inclusiveness*: for all initial constitutions C, we have $C \subseteq C^*$ – reasoning does not remove attitudes. *Monotonicity*: for all initial constitutions C and D, if $C \subseteq D$ then $C^* \subseteq D^*$ – additional attitudes cannot prevent (but can enable) new attitudes, equivalently additional absences of attitudes cannot enable (but can prevent) new attitudes.

does not rain. In convincing you, the argument has a causal effect on your attitudes: you gain the belief that it does not rain and lose the belief that it rains. By uttering the conclusion of the argument, you express both the belief acquisition and the belief loss.

Unlike Broomean reasoning, Druckerian reasoning is not explicit all the way. It is explicit in the sense that it follows an argument in language. But Drucker leaves the explicit paradigm in attributing to reasoning various implicit attitudinal changes that the argument induces. In the 'rain' example, the explicit reasoning by which you acquire the 'no rain' belief is both Druckerian and Broomean reasoning towards a belief, but the loss of the 'rain' belief is attributable to reasoning only in an implicit and non-Broomean sense.

Broome and Drucker both use the concept of 'expressing', but with different meanings. For Broome, your sentence expresses its (literal) content, in that it denotes or represents it. For Drucker, your utterance of a sentence expresses an attitude (change) of yours just if, according to the rules of the language, you could not utter that sentence sincerely while knowing that the utterance is not caused non-deviantly by the occurrence of the attitude (change).[25] Thus, for Broome, the conclusion sentence 'it does not rain' expresses the marked content of the 'no rain' belief. For Drucker, uttering that sentence expresses both the acquisition of the 'no rain' belief and the loss of the 'rain' belief.

In sum, Druckerian reasoning goes beyond Broomean reasoning in including processes that would count as automatic under our Broomean account. We find the Broomean notion of reasoning useful – for philosophy, but also cognitive science, decision theory, and behavioural science – because it aims at a clear conceptual separation between processes under a reasoner's explicit control and automatic processes beyond such control. This mirrors the psychological distinction between System 2 and System 1 processes (Watson and Evans 1974, Kahneman 2011). However, we acknowledge that the two kinds of process can interact in ways that are not represented explicitly in the Broomean model.[26]

5.3 Which logical requirements are achievable through reasoning?

What would it mean to achieve a logical requirement or even full rationality through reasoning? Given a theory of rationality, a reasoning system S **achieves** consistency, completeness, closedness, or (full) rationality if for each initial constitution $C \subseteq M$ the revision $C|S$ is, respectively, consistent, complete, closed, or rational.

[25] While Drucker only defines 'expressing an attitude' (p. 6), we read his definition as applying analogously to 'expressing attitude *changes*', because the latter is what is ultimately needed in his Argumentalism.

[26] Broome (2013: 206–207) points out that some automatic processes have semantic features, and that this fact raises 'interesting and difficult questions' that are outside the scope of his analysis.

We shall want reasoning to not only achieve certain requirements, but also to preserve consistency. Formally, a reasoning system S **preserves consistency** if for each consistent constitution C its revision $C|S$ is still consistent. Preserving consistency matters because there would be little point in achieving some requirement if one thereby lost consistency, the arguably most basic and 'least sacrificeable' logical requirement.

By Theorem 2, achieving consistency, completeness, or closedness is respectively equivalent to achieving certain rationality requirements of standard type. But whether these standard-type requirements are achievable is known; it is informally contained in Broome's work, and formally worked out in Dietrich et al. (2019). Details aside, reasoning can successfully achieve *closedness* requirements, but not consistency or completeness requirements. Using this fact, Theorem 2 implies another theorem as a corollary, which (roughly) says that

- reasoning can achieve closedness while preserving consistency,
- reasoning cannot achieve consistency,
- reasoning can achieve completeness, but only while sacrificing consistency.

Formally:

Theorem 3 *Given any theory of rationality,*

(a) some reasoning system achieves closedness while preserving consistency,

(b) no reasoning system achieves consistency (unless consistency is trivial[27]*),*

(c) no reasoning system achieves completeness while preserving consistency (unless completeness is essentially trivial[28]*),*

(d) no reasoning system achieves full rationality (unless consistency is trivial).

In (b)–(d), 'unless' can be read not only in its weak sense ('if it is not the case that'), but even in its strong sense ('if *and only if* it is not the case that'). So Theorem 3 provides necessary and sufficient conditions for the possibility of successful reasoning, in four senses of 'successful'.[29]

The message of Theorem 3 is gloomy, though 'Broomean': you cannot reason towards two of three logical requirements, just as (following Broome) you cannot reason towards many ordinary rationality requirements. This result is independent of the attitude type: it even holds for ordinary 'theoretic' reasoning in beliefs.

A more nuanced picture emerges after cashing in that other mental processes than reasoning could jump in to make your attitudes inch closer to completeness (by creating attitudes) or consistency (by removing attitudes). For instance, some beliefs or intentions might crowd out other ones that are inconsistent with them,

[27] i.e., unless the theory deems all constitutions consistent (or equivalently, deems the all-attitudes constitution $C = M$ rational).

[28] i.e., unless the theory deems essentially every constitution complete, in a sense defined below.

[29] In part (c), the stronger reading of 'unless' requires a compactness assumption: each inconsistent set of states $C \subseteq M$ has a finite inconsistent subset. Compactness holds trivially if M is finite. Compactness is the multi-attitude counterpart of ordinary logical compactness.

making you 'more consistent'. We *can* become 'more logical', but not through reasoning alone.

We now discuss each part in turn.

Part (a): the achievability of closedness. By part (a), you can develop closed attitudes through reasoning – without losing consistency. Why? By Theorem 2, closedness is achieved once all the theory's closedness *requirements* are achieved. A closedness requirement says: having a certain set of attitudes P implies having a certain attitude c. You achieve this requirement if you have the rule $r = (P, c)$. You achieve *all* of the theory's closedness requirements if you have *all* corresponding rules. If these are your only rules, reasoning provably preserves consistency. Although this reasoning system does the job, it is peculiar: it is so rich in rules that you can reason towards each closedness requirement of the theory in a single step. In practice, much slimmer (and cognitively more plausible) reasoning systems also achieve closedness and preserve consistency. You only need rules corresponding to *some* of the theory's closedness requirements. Suppose rationality requires that believing p and $if\ p\ then\ q$ implies believing q, and that believing q implies intending r. Then rationality also requires that believing p and $if\ p\ then\ q$ implies intending r. If you have the rules corresponding to the first two closedness requirements,

$$r = (\{(p, bel), (if\ p\ then\ q, bel)\}, (q, bel)) \text{ and } r' = (\{(q, bel)\}, (r, int)),$$

then you need not have the rule corresponding to the third requirement, $r'' = (\{(p, bel), (if\ p\ then\ q, bel)\}, (r, int))$, because the third requirement is achievable through applying first r and then r'. Real people presumably reason with few and simple rules.

Part (b): the unachievability of consistency. Part (b) is mathematically trivial, but philosophically disturbing. It is trivial (without even consulting Theorem 2) because Broomean reasoning never removes attitudes, hence never makes inconsistent constitutions consistent. Broome acknowledges that inconsistencies often disappear, but insists that they disappear, not through reasoning, but through automatic mental processes, such as when you find yourself losing a belief after realizing a conflict with other beliefs. The impossibility to *reason* yourself out of inconsistency is disturbing because consistency is a more basic normative desideratum than completeness and closedness. One would have hoped that reasoning can *at least* make consistent. Instead reasoning can make closed, but not consistent. The problem is only avoided for trivial theories of rationality that deem all constitutions consistent.

Part (c): the unachievability of completeness. Why does part (c) hold? Given the theory of rationality, we call a set of attitudes **avoidable** if some rational constitution contains none of its states, and **unavoidable** otherwise. Typical avoidable sets are $\{(p, bel), (not\ p, bel)\}$, $\{(p, int), (q, int), (r, int)\}$, and $\{(p, q, \succ$

$), (q, p, \succ), (p, q, \sim)\}$, for propositions p and q – though these sets are unavoidable if the theory requires holding 'beliefs about anything' and 'preferences between any options'. The theory's unavoidable sets stand in one-to-one correspondence with the theory's completeness requirements: a set $U \subseteq M$ is unavoidable if and only if the theory makes the completeness requirement of having some attitude from U. Now by Theorem 2, completeness is achieved once you satisfy the theory's completeness *requirements*, or equivalently, once you have acquired some attitude from each unavoidable set. There is a trivial (but implausible) way to acquire such attitudes: for each unavoidable set U, you simply have a rule that always generates a given attitude in U (formally, a rule $r = (\varnothing, m)$ which has no premise-attitudes and some conclusion-attitude m in U).

This trivial way to reason towards completeness is unconvincing. It seems ad hoc, if not stubborn and blind, to always acquire the same attitude from a given unavoidable set U, regardless of the web of existing attitudes. What matters is not just *that* you form an intention (from an unavoidable set of intentions U), but also *which* intention you form. Otherwise the new intention can be inconsistent with your beliefs, preferences, or other existing attitudes. Formally, the trivial reasoning system achieves completeness by sacrificing consistency. Unfortunately, also all other reasoning systems that achieve completeness fail to preserve consistency.

This argument presupposes that completeness is not essentially trivial, as shown in the proof. Completeness is **trivial** if the theory deems all constitutions complete; or equivalently, the empty constitution is rational. Here there are no unavoidable sets U. More generally, completeness is **essentially trivial** if all constitutions *containing at least the unfalsifiable attitudes (if any)* are complete; or equivalently, some constitution containing at most unfalsifiable attitudes is rational. An attitude m is **unfalsifiable** if it never conflicts with other attitudes, i.e., if $\{m\} \cup C$ is consistent whenever C is consistent. Standard theories of rationality deem no attitudes unfalsifiable: desiring p is falsifiable by conflicting with desiring *not* p; preferring p to q is falsifiable by conflicting with being indifferent between p and q, or with preferring q to p, or with preferring q to r and also r to p; etc. So, for standard theories of rationality, essentially trivial completeness just means trivial completeness.

Part (d): the unachievability of full rationality. Since consistency is unachievable by part (b), so is full rationality. This again presupposes that not all constitutions count as consistent – otherwise you could trivially become rational by having *all* reasoning rules, making you form all attitudes.

6 Abstract rationality versus concrete logics of rational attitudes

Our abstract model of multi-attitude psychology employs no concrete logic, i.e., no formal syntax or semantics. There exist many concrete logics of attitudes, such as beliefs or preferences. This section briefly compares our abstract approach with concrete logical approaches, at the level of the statics of attitudes (Section 6.1) and the dynamics of attitudes (Section 6.2).

6.1 The statics of multiple attitudes

The statics of multi-attitude psychology concern your attitudes *at a given time*. Our abstract logical requirements – consistency, completeness, closedness – are purely static. An alternative to our abstract approach would be to use some concrete logic of attitudes. Mono-modal logics involve just one attitude, for instance belief in 'doxastic logics' (e.g., Halpern 2005) or preferences in 'preference logics' (e.g., Liu 2011). Multi-modal logics involve two or more attitudes, for instance beliefs, desires and intentions in 'BDI logics' (e.g., Van der Hoek and Wooldridge 2003). Attitudes are represented by modal operators, and rationality by axioms that constrain attitudes. This machinery provides concrete representations of attitude types (through attitude operators), but also of attitude contents (through logical sentences). Since these contents can themselves involve attitudes, one can explicitly construct and study nested attitudes (meta-attitudes) such as intentions to desire to believe something. Like our abstract model, such a concrete logical model can of course be used to define notions of attitudinal consistency, completeness, and closedness, though one would be limited to the (often few) attitudes present in the logic in question.

6.2 The dynamics of multiple attitudes

The dynamics of multi-attitude psychology concern attitude *change*. Modal logics of the sort just discussed can model (deductive[30]) reasoning about attitudes, through the entailment relation. But reasoning about attitudes is a process of attitude discovery, not attitude change; it is not reasoning *in* attitudes (or *with* attitudes, in Broome's words). Establishing that this difference is real and could not be easily overcome through some formal reduction requires a careful analysis, which we undertake in Dietrich and Staras (2022). Here, a few remarks should suffice. If someone reasons about your attitudes, then what changes are not your attitudes, but the reasoner's beliefs about them. Even when it is you yourself who reason

[30]But logical entailment cannot model *non-deductive* reasoning. According to the dominant view in philosophy of logic, crystallised by Harman's (1984) distinction between inference and implication, logic is not primarily about reasoning (inference), but about entailment (implication). Christensen (2004) also analyses this distinction.

about some of your attitudes, then not *those* attitudes change, but your (meta-)beliefs about them.[31] In our earlier example, you reason *in* your attitudes by saying:

Doctors recommend resting. So, I shall rest.

You thereby form an intention from a belief. An observer (possibly you) might reason *about* your attitudes by saying:

You believe doctors recommend resting. So, you intend to rest.

This and other reasoning about attitudes can be modelled modal-logically, using entailments between atomic attitude-sentences of type 'you hold attitude such-and-such towards such-and-such', formally $O(\phi)$ with an operator O representing the attitude type and a sentence ϕ representing the attitude content. Thanks to building appropriate rationality axioms into the logic, the right entailments hold between such 'atomic' attitude-sentences. The logic also provides entailments between plenty of 'non-atomic' attitude-sentences, such as: 'you do *not* desire this', 'you *either* believe this *or* intend that', etc. Reasoning about attitudes can thus start from, or conclude in absences of attitudes (or disjunctions of attitudes etc.) – meaning that the reasoner discovers that absence (or disjunction etc.). But Broomean reasoning *in* attitudes cannot start from, or conclude in, absences – meaning that reasoning starts from attitudes you *have* and generates rather than removes attitudes (cf. Section 5.2).

In sum, analogies between our abstract approach to multi-attitude psychology and concrete attitude logics are easier to draw at the static level than at the dynamic level. At the static level, both approaches include notions of consistency, completeness and closedness. At the dynamic level, one should conceptually distinguish our abstract model of reasoning from entailment in concrete attitude logics, as the former captures reasoning in attitudes while the latter captures reasoning about attitudes.

A Relation between our logical conditions on attitudes and standard logical conditions on beliefs

This appendix clarifies how our three logical conditions on multiple attitudes generalise standard logical conditions on beliefs only. We continue to assume that our logical conditions are derived from a theory of rationality, i.e., we retain the top-down approach of Section 3.1 that has guided much of our analysis.

Informally, the standard logical conditions on (a set of) beliefs say the following:

[31] Your introspective reasoning may spark some causal process that changes your attitudes (in some direction), but this is another issue.

(a) *Consistency* says: believe only propositions that are mutually consistent, i.e., can be simultaneously true.

(b) *Completeness* comes in two variants. *Local* completeness says: believe a member of each proposition-negation pair $\{p, \textit{not } p\}$. *General or global* completeness says something stronger: believe a member of each set of mutually exhaustive propositions, i.e., propositions that cannot be simultaneously false. There are many such sets: proposition-negation pairs $\{p, \textit{not } p\}$, sets of type $\{p, q, [\textit{not } p] \textit{ or } [\textit{not } q]\}$, etc.

(c) *Closedness* says: believe all deductive consequences of your beliefs, i.e., all beliefs that are true whenever your existing beliefs are true.

To state these definitions formally, consider a set L of *propositions* defined syntactically or semantically, as in the 'illustration' in Section 2.[32] A *belief set* is a set of ('believed') propositions $B \subseteq L$. It is

- *consistent* if it is members can be jointly true. Given the semantic model, this means that $\cap_{b \in B} b \neq \varnothing$. Given the syntactic model, it means that B entails no contradiction.
- *closed* if it contains all $p \in L$ which it entails. In the semantic model, B entails p just in case $\cap_{b \in B} b \subseteq p$.
- *locally complete* if it contains a member of each proposition-negation pair, i.e., each pair $\{p, \Omega \backslash p\} \subseteq L$ (given the semantic model) or each pair $\{p, \neg p\} \subseteq L$ (given the syntactic model).
- *globally complete* if it contains a member of each exhaustive set $Y \subseteq L$. A set $Y \subseteq L$ is *exhaustive* if necessarily at least one member is true. i.e., if $\cup_{p \in Y} p = \Omega$ (given the semantic model) or if the set $\{\neg p : p \in Y\}$ is inconsistent (given the syntactic model). The simplest exhaustive sets are the proposition-negation pairs. Global completeness implies local completeness, since local completeness quantifies over fewer exhaustive sets, namely only over proposition-negation pairs. An equivalent definition of 'globally complete' is given in Lemma 1(b).

We can now compare these standard conditions to ours.

A difference between our and standard logical conditions. While our logical conditions on attitudes are derived from a notion of rationality and are thus by definition requirements of rationality, the standard logical conditions on beliefs may or not be required, depending on what counts as rational for beliefs. Completeness is controversial as a rationality requirement on beliefs, while consistency and closedness are widely accepted. We shall therefore regard a belief set $B \subseteq L$

[32] In the syntactic case we assume that the logic is a standard propositional logic, or more generally any well-behaved logic such as a standard propositional, predicate, modal, or conditional logic. Formally, the logic must obey a few classic conditions (namely L1–L4 in Dietrich 2007) which guarantee 'regular' notions of logical consistency and logical entailment. The notable condition is monotonicity, whereby entailments are preserved under adding premises, and so consistency of a set is preserved under removing elements.

as rational in the standard sense if it is consistent and closed, and as rational in a stronger sense if it is moreover complete (in the local or global sense, which are equivalent given consistency).

The conditional equivalence between our and standard logical conditions. Our logical conditions on attitudes are equivalent to the ordinary logical conditions on beliefs if beliefs are the only attitudes and the theory of rationality is standard. Why? We assume the framework of the 'illustration' in Section 2 in the belief-only special case $A = \{bel\}$ (where L contains semantic or syntactic propositions[33]). So M contains only belief-attitudes: $M = \{(p, bel) : p \in L\}$. Beliefs being the only attitudes, constitutions are equivalent to belief sets: to any constitution $C \subseteq M$ corresponds a belief set $B = \{p \in L : (p, bel) \in C\}$, and to any belief set $B \subseteq L$ corresponds a constitution $C = \{(p, bel) : p \in B\}$. In this belief-only framework, theories of rationality T are essentially theories of rational *beliefs*. Two theories of rationality are particularly salient in this context, as they reflect what is usually required from beliefs:

- The *standard* theory of rationality is the theory T_{stan} such that a constitution $C \subseteq M$ is rational (i.e., in T_{stan}) if and only if the corresponding belief set $B = \{p : (p, bel) \in C\}$ is consistent and closed.
- The *standard complete* theory of rationality is the theory $T_{\text{stan+}}$ such that a constitution $C \subseteq M$ is rational (i.e., in $T_{\text{stan+}}$) if and only if the corresponding belief set $B = \{p : (p, bel) \in C\}$ is consistent and complete (in the local or, equivalently, global sense), and thus by implication closed. Note that $T_{\text{stan+}} \subseteq T_{\text{stan}}$.

Our logical conditions then reduce to the standard ones:

Theorem 4 *In the above belief-only framework, for any constitution C with corresponding belief set B,*

(a) C is consistent under theory $T_{\text{stan+}}$ or T_{stan} if and only if B is consistent,

(b) C is complete under theory $T_{\text{stan+}}$ if and only if B is complete (understood globally),

(c) C is closed under theory $T_{\text{stan+}}$ or T_{stan} if and only if B is closed.

This result precisifies how our logical conditions generalise the ordinary ones. The connection is tight for consistency and closedness, and weaker for completeness, reinforcing arguments in Section 3.4.

To prove Theorem 4, we start by characterizing the standard logical conditions on beliefs in ways similar to our definition of logical conditions on constitutions:

Lemma 1 *A belief set $B \subseteq L$ is*

(a) consistent if and only if $B \subseteq B'$ for some complete and consistent belief set $B' \subseteq L$,

[33] and where in the syntactic case the logic is well-behaved as defined in footnote 32

(b) complete (understood globally) if and only if $B \supseteq B'$ for some complete and consistent belief set B',

(c) closed if and only if B contains each proposition which it entails, where being entailed means being contained in all complete and consistent extensions $B' \supseteq B$.

Proof. Let $B \subseteq L$ be a belief set, and $\mathbf{B}$ the set of complete and consistent belief sets.

(a) We distinguish between the semantic and syntactic model of L. In the semantic case the equivalence holds trivially (if B is consistent, we can pick a $w \in \cap_{p \in B} p$ and define B' as $\{p \in L : w \in p\}$). In the syntactic case the equivalence follows from a basic property in logic, often referred to as 'Lindenbaum's lemma', which states that any consistent set of sentences in a logic is extendable to a complete and still consistent set. This property holds in well-behaved logics of the sort assumed here (see footnote 32).

(b) First let B have a subset $B' \in \mathbf{B}$. To show that B is (globally) complete, consider any exhaustive set $Y \subseteq L$. We must prove that $B \cap Y \neq \varnothing$. As $B' \subseteq B$ it suffices to show that $Y \cap B' \neq \varnothing$, which holds by the following argument, spelt out separately for syntactic and semantic propositions:

- *In the syntactic case*, note that the (inconsistent) set $\{\neg p : p \in Y\}$ cannot be a subset of the (consistent) set B'. So there is a $p \in Y$ such that $\neg p \notin B'$, and thus $p \in B'$ as B' is complete. So $Y \cap B' \neq \varnothing$.
- *In the semantic case*, since $\{\Omega \backslash p : p \in Y\}$ has empty intersection (as Y has union Ω) while B' has non-empty intersection (as B' is consistent), the set $\{\Omega \backslash p : p \in Y\}$ cannot be a subset of B'. So there is a $p \in Y$ such that $\Omega \backslash p \notin B'$, and hence $p \in B'$ as B' is complete. So $Y \cap B' \neq \varnothing$.

Conversely, assume that B does *not* include any $B' \in \mathbf{B}$. We show that B is not (globally) complete. By assumption, for each $B' \in \mathbf{B}$ there is a $p_{B'} \in B' \backslash B$. Let $Y := \{p_{B'} : B' \in \mathbf{B}\}$. This set Y is exhaustive – in the semantic case because each world $\omega \in \Omega$ belongs to some member of Y (namely to $p_{B'}$ where $B' := \{p \in L : \omega \in p\}$), in the syntactic case because $\{\neg p : p \in Y\}$ is not included in any $B' \in \mathbf{B}$ and so is inconsistent by (a). Yet $Y \cap B = \varnothing$ by construction of Y. So B is not complete.

(c) We show that B is closed just in case $B = \cap_{B' \in \mathbf{B}: B' \supseteq B} B'$. In the syntactic case, this is a familiar fact (in the well-behaved logics considered here; cf. footnote 32). Now consider the semantic case. If $B = \cap_{B' \in \mathbf{B}: B' \supseteq B} B'$, then B is clearly closed. Conversely, if B is closed, then $B = \{p \in L : p \supseteq \cap_{q \in B} q\}$, implying $B = \cap_{B' \in \mathbf{B}: B' \supseteq B} B'$. ■

Proof of Theorem 4. Suppose the theorem's assumptions. For brevity, we only prove the claims relating to the theory $T_{\text{stan+}}$. Write T for $T_{\text{stan+}}$. Denote the content of a (belief) state m by $\widehat{m}$ and the belief set corresponding to a constitution $C \subseteq M$ by $\widehat{C} = \{\widehat{m} : m \in C\}$. Fix a constitution C.

First,

$$\begin{aligned} C \text{ is consistent} &\Leftrightarrow C \subseteq C' \text{ for some } C' \in T \\ &\Leftrightarrow \widehat{C} \subseteq \widehat{C'} \text{ for some } C' \in T \\ &\Leftrightarrow \widehat{C} \subseteq B \text{ for some consistent and complete } B \subseteq L \\ &\Leftrightarrow \widehat{C} \text{ is consistent, by Lemma 1(a).} \end{aligned}$$

Second,

$$\begin{aligned} C \text{ is complete} &\Leftrightarrow C \supseteq C' \text{ for some } C' \in T \\ &\Leftrightarrow \widehat{C} \supseteq \widehat{C'} \text{ for some } C' \in T \\ &\Leftrightarrow \widehat{C} \supseteq B \text{ for some consistent and complete } B \subseteq L \\ &\Leftrightarrow \widehat{C} \text{ is globally complete, by Lemma 1(b).} \end{aligned}$$

Third, writing $\widehat{T} := \{\widehat{C} : C \in T\} = \{B \subseteq L : B \text{ is complete and consistent}\}$,

$$\begin{aligned} C \text{ is closed} &\Leftrightarrow C \ni m \text{ for all } m \text{ entailed by } C, \text{ i.e., all } m \in \cap_{C' \in T : C' \supseteq C} C' \\ &\Leftrightarrow \widehat{C} \ni \widehat{m} \text{ for all } m \text{ entailed by } C, \text{ i.e., all } m \in \cap_{C' \in T : C' \supseteq C} C' \\ &\Leftrightarrow \widehat{C} \ni b \text{ for all } b \text{ entailed by } \widehat{C}, \text{ i.e., all } b \in \cap_{B \in \widehat{T} : B \supseteq \widehat{C}} B \\ &\Leftrightarrow \widehat{C} \text{ is closed, by Lemma 1(c).} \blacksquare \end{aligned}$$

B Proof of Theorem 1 and its corollary

The proof of Theorem 1 rests on two lemmas. The first lemma is an equivalent re-statement of a well-known fact in abstract logic, whose proof we include for completeness:

Lemma 2 *A set CLO ($\subseteq 2^M$) is a closedness notion if and only if it is closed under intersection, i.e., $Y \subseteq CLO \Rightarrow \cap Y \in CLO$ (where by convention $\cap\varnothing = M$).*

Proof. First, if CLO is closed under intersection, then define the consequence $Cn(C)$ of a set $C \subseteq M$ as the smallest extension of C in CLO, i.e., as $\cap\{C' \in CLO : C' \supseteq C\}$; and verify that the so-defined operator Cn is classical and that $CLO = \{C \subseteq M : Cn(C) = C\}$. Conversely, assume CLO is a closedness notion, say with respect to the classical consequence operator Cn. To show closedness under intersection, we fix a $Y \subseteq CLO$ and show that $\cap Y \in CLO$, i.e., that $Cn(\cap Y) = \cap Y$. For one, $\cap Y \subseteq Cn(\cap Y)$, as Cn is inclusive. For another, $Cn(\cap Y) \subseteq \cap Y$, as for each $C \in Y$ we have $Cn(\cap Y) \subseteq Cn(C) = C$, where the '$\subseteq$' holds as Cn is monotonic and the '$=$' holds as $C \in CLO$. $\blacksquare$

Lemma 3 *For any theory of rationality T, CLO_T is the closure of T under intersection, i.e., $CLO_T = \{\cap Y : Y \subseteq T\}$.*

Proof. Let T be any theory. Since CLO_T includes T and is, like any closedness notion, closed under intersection (Lemma **??**), CLO_T includes T's closure under

intersection: $CLO_T \supseteq \{\cap Y : Y \subseteq T\}$. To show that $CLO_T \subseteq \{\cap Y : Y \subseteq T\}$, we fix a $C \in CLO_T$, define $Y = \{C' \in T : C \subseteq C'\}$, and show that $C = \cap Y$. All attitudes in $\cap Y$ are (T-)entailed by C, hence belong to C as C is (T-)closed. So, $C = \cap Y$. ■

Proof of Theorem 1. Assume T is classical, say $T = CON \cap COM \cap CLO$ for some notions of consistency CON, completeness COM, and closedness CLO. We must show that (i) $T = CON_T \cap COM_T \cap CLO_T$, (ii) $CON_T \subseteq CON$, (iii) $COM_T \subseteq COM$ and (iv) $CLO_T \subseteq CLO$. Observe that (i) follows from (ii)–(iv) and the fact that $T = CON \cap COM \cap CLO$, $T \subseteq CON_T$, $T \subseteq COM_T$, and $T \subseteq CLO_T$. So it suffices to show (ii)–(iv).

To show (ii), let $C \in CON_T$. Pick a $C' \in T$ such that $C \subseteq C'$. As $C' \in T$ and $T \subseteq CON$, we have $C' \in CON$; hence $C \in CON$ since CON is a consistency notion and $C \subseteq C'$. This shows (ii).

To show (iii), consider a $C \in COM_T$. Pick a $C' \in T$ such that $C' \subseteq C$. Now $C' \in COM$ (as $C' \in T \subseteq COM$), and thus $C \in COM$ (as $C' \subseteq C$ and COM is a completeness notion).

We finally show (iv). As CLO includes T and is closed under intersection (by Lemma 2), CLO includes T's closure under intersection, which equals CLO_T by Lemma 3. So, $CLO \supseteq CLO_T$. ■

Proof of Corollary 1. Let T be fully classical, say $T = CON \cap CLO$ for notions of consistency CON and closedness CLO. Then (*) $T = CON \cap COM \cap CLO$ with the vacuous completeness notion $COM = 2^M$. By Theorem 1, (**) $T = CON_T \cap COM_T \cap CLO_T$, and (***) $CON_T \subseteq CON$, $COM_T \subseteq COM$, $CLO_T \subseteq CLO$. It remains to show that $T = CON_T \cap CLO_T$. By (*), (**) and (***), $T = CON_T \cap COM \cap CLO_T$. So, as $COM = 2^M$, $T = CON_T \cap CLO_T$. ■

We also present an (alternative) direct proof of Corollary 1. It is again based on two lemmas, namely on Lemma 2 and on an interesting fact about fully classical theories:

Lemma 4 *If T is a fully classical theory of rationality, then $CLO_T = T \cup \{M\}$.*

Proof. Obviously, for any theory T (whether or not fully classical) CLO_T includes T and contains M; hence, $T \cup \{M\} \subseteq CLO_T$. To show the reverse inclusion, let T be fully classical, say $T = CON \cap CLO$ for notions of consistency CON and closedness CLO. By Lemma 2, CLO is closed under intersection: $Y \subseteq CLO \Rightarrow \cap Y \in CLO$. Moreover, CON is closed under *non-empty* intersection: $\varnothing \neq Y \subseteq CON \Rightarrow \cap Y \in CON$. It follows that $T = CON \cap CLO$ is closed under *non-empty* intersection. By implication, $T \cup \{M\}$ is closed under intersection.

We are ready to show that $CLO_T \subseteq T \cup \{M\}$. We let $C \in CLO_T$ and prove $C \in T \cup \{M\}$. Now C (T-)entails all attitudes in $\cap\{C' \in T : C \subseteq C'\}$, hence contains all of them as C is (T-)closed. So, $C = \cap\{C' \in T : C \subseteq C'\}$. Meanwhile

$\cap\{C' \in T : C \subseteq C'\} \in T \cup \{M\}$ as $T \cup \{M\}$ is closed under intersection. Therefore $C \in T \cup \{M\}$. ■

Direct proof of Corollary 1. Suppose T is a fully classical theory, say $T = CON \cap CLO$ for notions of consistency CON and closedness CLO. We must show that (i) $T = CON_T \cap CLO_T$, (ii) $CON_T \subseteq CON$, and (iii) $CLO_T \subseteq CLO$. Note that (i) follows from (ii) and (iii) because $T = CON \cap CLO$, $T \subseteq CON_T$, and $T \subseteq CLO_T$. It thus remains to show (ii) and (iii). Condition (ii) holds for the same reason as in the proof of Theorem 1. To show (iii), we must by Lemma 4 prove that $T \cup \{M\} \subseteq CLO$. This holds because $T \subseteq CLO$ (as $T = CON \cap CLO$) and because $M \in CLO$ by Lemma 2. ■

C Proof of Theorem 2

Fix a theory of rationality T and a constitution C. Let $T \neq \varnothing$, an assumption needed only in parts (a) and (b). We now prove each part.

Part (a). We prove both directions of implication. We may assume $C \neq \varnothing$, since otherwise C is trivially consistent (as $T \neq \varnothing$) and satisfies all consistency requirements.

- First let C satisfy T's consistency requirements. We show that C is consistent. Consider the consistency requirement R^* of not holding all states in C: $R^* = \{C' : C \not\subseteq C'\}$. Since C violates R^* while satisfying T's consistency requirements, R^* cannot be a requirement of T. So some rational constitution $C' \in T$ violates R^*, i.e., $C \subseteq C'$. So C is consistent.
- Conversely, assume C is consistent. Consider any consistency requirement R of T; we must prove that C satisfies it. R takes the form $R = \{C' : F \not\subseteq C'\}$ for some 'forbidden set' F. Being consistent, C has a rational extension C^+. As C^+ is rational, it satisfies T's requirements, so satisfies R, i.e., $F \not\subseteq C^+$. As $C \subseteq C^+$, it follows that $F \not\subseteq C$. So C satisfies R.

Part (b). The proof is the 'dual' of that for part (a). We may suppose $C \neq M$, because otherwise C is trivially complete (as $T \neq \varnothing$) and satisfies all completeness requirements.

- First let C satisfy T's completeness requirements. We show that C is complete. Note that C violates the (completeness) requirement of containing a state outside C, $R^* = \{C' : (M \backslash C) \cap C' \neq \varnothing\}$. So, as C satisfies T's completeness requirements, R^* is not a requirement of T. So some rational constitution $C' \in T$ violates R^*; hence $(M \backslash C) \cap C' = \varnothing$, i.e., $C' \subseteq C$. So C is complete.
- Conversely, let C be complete. Let R be any completeness requirement of T; we show that C satisfies it. R requires having at least one states from an (unavoidable) set U: $R = \{C' : C' \cap U \neq \varnothing\}$. As C is complete, it has a rational subset C^-. Being rational, C^- satisfies T's requirements, hence

satisfies R, i.e., $C^- \cap U \neq \varnothing$. So, as $C^- \subseteq C$, $C \cap U \neq \varnothing$. Hence, C satisfies R.

Part (c). Again, both directions of implication are to be shown.

- First, let C satisfy T's closedness requirements. To show that C is closed, consider a state m entailed by C; we must show that $m \in C$. Consider the closedness requirement R^* with set of premise states C and conclusion state m: $R^* = \{C' : C \subseteq C' \Rightarrow m \in C'\}$. As C entails m, R^* is a requirement of T. So, as C satisfies T's closedness requirements, it satisfies R^*. Hence, as $C \subseteq C$, we have $m \in C$.
- Conversely, assume C is closed. Consider a closedness requirement R of the theory, say $R = \{C' : P \subseteq C' \Rightarrow c \in C'\}$ for some (premise) set $P \subseteq M$ and some (conclusion) state $c \in M$. To show that C satisfies R, assume $P \subseteq C$; we must prove $c \in C$. Since R is a requirement of T, all rational constitutions which include P contain c, which in turn means that P entails c (by definition of entailment). So also the larger set $C \supseteq P$ entails c (again by definition of entailment). Hence $c \in C$, as C is closed.

Part (d). Trivially, rationality is equivalent to satisfaction of the theory's strongest requirement $R = T$, which is equivalent to satisfaction of all the theory's requirements $R \supseteq T$. ■

D Proof of Theorem 3

Again, fix a theory of rationality T. A reasoning system S **achieves** a requirement R if $C|S$ satisfies R for all constitutions C. For parts (b), (c) and (d) we prove two directions of implication, as 'unless' is taken to mean 'if *and only if* it is not the case that'.

For the trivial theory $T = \varnothing$, all parts hold. Part (a) holds because the maximal reasoning system S, which contains all rules, achieves closedness (by transforming each constitution into M, the only closed constitution) and trivially preserves consistency since no constitution is consistent. Parts (b), (c) and (d) hold because consistency, completeness and rationality are all trivially unachievable by the absence of any consistent, complete or rational constitutions (regarding (c), note also the absence of avoidable sets).

Henceforth let $T \neq \varnothing$. We prove the four parts in turn.

Part (a). By Theorem 2(c), achieving closedness is equivalent to achieving all closedness requirements of T. Meanwhile, by Theorem 1 in Dietrich et al. (2019) there exists a reasoning schema S which achieves all closedness requirements and preserves consistency. So S achieves closedness while preserving consistency.

Part (b). First, if consistency is trivial (i.e., $C = M$ is rational), then consistency is achieved by any reasoning system. Conversely, assume consistency is non-trivial.

Let S be any reasoning system. It fails to achieve consistency, because by non-triviality there is an inconsistent constitution C (e.g., $C = M$), and as $C|S \supseteq C$ also $C|S$ is inconsistent.

Part (c). First, assume completeness is trivial (along with the background assumption of compactness, whereby each inconsistent set of states has a finite inconsistent subset). For each unavoidable set U we can pick an unfalsifiable state $m_U \in U$. The reasoning system $S = \{(\varnothing, m_U) : U \text{ is unavoidable}\}$ achieves each completeness requirement of theory T, because for each completeness requirement of T a state from its unavoidable set is formed. So S achieves completeness simpliciter, by Theorem 2. We now show that S preserves consistency. For a contradiction, consider a consistent constitution C such that $C|S$ is inconsistent. By compactness, $C|S$ has a finite inconsistent subset C'. By definition of S, $C|S = C \cup \{m_U : U \text{ is an unavoidable set}\}$. So we may pick finitely many unavoidable sets $U_1, ..., U_k$ such that $C' \subseteq C \cup \{m_{U_1}, m_{U_2}, ..., m_{U_k}\}$. Since C is consistent, so is $C \cup \{m_{U_1}\}$, as m_{U_1} is non-falsifiable; hence so is $C \cup \{m_{U_1}, m_{U_2}\}$, as m_{U_2} is non-falsifiable. Repeating this argument k times, it follows that $C \cup \{m_{U_1}, m_{U_2}, ..., m_{U_k}\}$ is consistent. Hence its subset C' is consistent.

Conversely, suppose some set of falsifiable states is unavoidable. Let R be the corresponding completeness requirement. It suffices to show that no reasoning system achieves R, because by Theorem 2 achieving completeness is equivalent to achieving all completeness requirements of the theory. By Theorem 3 in Dietrich et al. (2019), no reasoning system achieves any completeness requirement of the theory whose unavoidable set consists of falsifiable states. So no reasoning system S achieves R.

Part (d). First, for (degenerate) theories that deem $C = M$ rational, rationality is trivially achieved by the reasoning system S containing *all* rules, for which $C|S = M$ for all initial constitutions C. Conversely, if $C = M$ is irrational, the unachievability of rationality follows from that of the weaker demand of consistency (see part (b)). ■

References

Alchourrón, Carlos E., Gärdenfors, Peter, Makinson, David (1985) On the logic of theory change: Partial meet contraction and revision functions, *The Journal of Symbolic Logic* 50 (2), 510-530

Boghossian, Paul (2014) What is Reasoning? *Philosophical Studies* 169: 1–18

Boghossian, Paul (2018) Delimiting the Boundaries of Inference. *Philosophical Issues* 28: 55–69

Broome, John (2007) Wide or narrow scope? *Mind* 116: 360-370

Broome, John (2013) *Rationality through reasoning*, Hoboken: Wiley

Christensen, David (2004) *Putting Logic in its Place: Formal Constraints on Rational Belief*, Oxford University Press, New York

Dietrich, Franz (2007) A generalised model of judgment aggregation, *Social Choice and Welfare* 28(4): 529-565

Dietrich, Franz, Christian List (2016) Mentalism versus behaviourism in economics: a philosophy-of-science perspective, *Economics and Philosophy* 32(3): 249-281

Dietrich, Franz, Antonios Staras (2022) Reasoning in versus about attitudes: Forming versus discovering one's mental states, working paper

Dietrich, Franz, Antonios Staras, Robert Sugden (2019) A Broomean model of rationality and reasoning, *Journal of Philosophy* 116: 585-614

Drucker, Daniel (2021) Reasoning beyond belief acquisition, *Noûs*, see https://doi.org/10.1111/no

Halpern, Joseph Y. (2005) *Reasoning About Uncertainty*, Cambridge, Massachusetts & London: MIT Press

Harman, Gilbert (1984), Logic and reasoning, *Synthese* 60(1):107-127

Horty, John F. (2001) Nonmonotonic Logic, in Goble, Lou, ed., *The Blackwell Guide to Philosophical Logic*, Blackwell

Kahneman, Daniel (2011) *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux

King, Jeffrey C. (2019) Structured Propositions, *The Stanford Encyclopedia of Philosophy*, Summer 2019 Edition

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

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

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

Tarski, Alfred (1956) *Logic, semantics and metamathematics*, Oxford University Press

Van der Hoek, Wiebe, Michael Wooldridge (2003) Towards a logic of rational agency, *Logic Journal of IGPL* 11(2): 135-159

Watson, Peter C., Evans, Jonathan B. (1974) Dual processes in reasoning?, *Cognition* 3(2): 141–54