

# THE ESSENCE OF STRUCTURAL IRRATIONALITY

## THE IMPOSSIBILITY OF ATTITUDINAL SUCCESS

*Julian Fink*

IT IS COMMONLY ACCEPTED that there is a wide range of combinations of attitudes that make a person irrational. These so-called structural cases of attitude-based irrationality include contradictory beliefs and intentions, failing to intend something you deem necessary for your intended ends (instrumental incoherence), failing to intend what you judge you ought to do (akratic incoherence), preferring *a* to *b*, *b* to *c*, but also *c* to *a* (cyclical preferences), etc.

Until recently, it was only tentatively assumed that these diverse patterns of attitudes belong to one and the same domain of irrationality. The tentative nature of this assumption has been challenged, however, by two major attempts to unify the domain of structural irrationality. Errol Lord and Benjamin Kiesewetter both propose that structural irrationality is a matter of combining attitudes that jointly guarantee the violation of a decisive normative reason.<sup>1</sup> Alex Worsnip, by contrast, suggests that structural irrationality is a matter of combining attitudes a person is constitutively disposed not to combine.<sup>2</sup>

Lord and Kiesewetter's proposal comes in two parts. First, they suggest that one is *substantially* irrational whenever one fails to respond correctly to the normative reasons that are epistemically available. For example, suppose you believe that the cat is on the mat, despite possessing sufficient evidence that this is not the case. Then, assuming that sufficient evidence translates to having decisive reason, you believe something you have decisive reason not to believe.<sup>3</sup> You therefore fail to respond correctly to the reasons available. This makes you substantially irrational.<sup>4</sup>

1 Lord, "What You're Rationally Required to Do and What You Ought to Do" and *The Importance of Being Rational*; Kiesewetter, *The Normativity of Rationality*.

2 Worsnip, "What Is (In)Coherence?" and *Fitting Things Together*.

3 On the point of sufficient evidence translating to having decisive reason, see Kiesewetter, *The Normativity of Rationality*, 180–85.

4 Cf. Kiesewetter, *The Normativity of Rationality*, ch. 7.

Second, *structural* irrationality is then understood as a subdomain of substantial irrationality. What creates this distinctive domain is the hypothesis that structurally irrational patterns of attitudes alone suffice to guarantee a failure to respond correctly to available reasons.<sup>5</sup>

By contrast, Worsnip offers a naturalistic alternative to the reasons-based account of structural irrationality.<sup>6</sup> Put roughly, structural irrationality resides in combinations of attitudes that a person is disposed not to sustain. More precisely, two or more attitudes are structurally irrational if and only if you are necessarily (and appropriately) disposed to abandon at least one of them once you become aware of holding them together.<sup>7</sup>

No doubt, both proposals have plenty of merit. They pick out aspects that are deeply symptomatic of many instances of structural irrationality. I am convinced that, generally speaking, structural irrationality tends to be unsustainable under awareness. This is particularly true of paradigmatic cases such as contradictory beliefs and intentions and means-end irrationality.

Likewise, there are paradigmatic instances of structural irrationality where you will necessarily violate a decisive reason. It is quite plausible, for example, that if you have sufficient evidence that you ought to *p*, then you will have decisive reason to intend to *p*. Also, if you lack sufficient evidence that you ought to *p*, then you have decisive reason not to believe that you ought to *p*. And so, since, necessarily, you either have or lack sufficient evidence that you ought to *p*, you indeed inevitably violate a decisive reason whenever you believe that you ought to do something without intending to do it.<sup>8</sup>

5 Kiesewetter, *The Normativity of Rationality*, 236; Lord, *The Importance of Being Rational*, 27. For example, suppose you believe *p* and you believe not-*p*. This is structurally irrational because, necessarily, either (you have decisive reason not to believe *p*) or (you have decisive reason not to believe not-*p*). More generally: whenever you adopt a structurally irrational pattern of attitudes, you necessarily have at least one attitude that is substantially irrational (i.e., you have decisive reason not to have it). It is this necessity that distinguishes structural irrationality from other (and more general) forms of irrationality.

6 Worsnip, "What Is (In)Coherence?" and *Fitting Things Together*.

7 Worsnip argues that the disposition to abandon an attitude must be sensitive to a *constitutive* aspect of the attitude in question. This is what I mean by "appropriately disposed": "That is, human agents are disposed such that they are (at least normally) not able to (or at least find it difficult to) psychologically sustain such combinations of attitudes under conditions of full transparency" (Worsnip, "What Is (In)Coherence?" 188). For example, simultaneously intending *p* and intending not-*p* is structurally irrational because, necessarily, once you become aware of having these two intentions, you are appropriately disposed either to give up your intention to *p* or give up your intention to not-*p*. Accordingly, structural irrationality consists of combinations of attitudes that tend not to survive cognitive transparency.

8 Cf. Kiesewetter, *The Normativity of Rationality*, 233 and sec. 9.5; Lord, *The Importance of Being Rational*, sec. 2.4.5.

Nevertheless, and despite these merits, I argue that both accounts fail to identify the *essence* of structural irrationality. Worsnip himself admits, for instance, that akratic incoherence (i.e., failing to intend what you believe you ought to do) poses a hard case for his account, since there is “widespread consensus that clear-eyed akrasia is possible.”<sup>9</sup> Thus, the inability (or even the disposition not) to sustain structural instances of irrationality under awareness does not seem to be a strictly necessary condition for structural irrationality.<sup>10</sup>

Moreover, there are also cases where the inability to sustain a combination of attitudes under transparency does not seem to be sufficient for identifying structural irrationality. Suppose you have attitude *A*, yet you lack a belief that you have attitude *A*. For many types of attitudes, this clearly fails to be structurally irrational. If, for instance, you desire to go skiing yet you lack a belief that you desire to go skiing, you are not necessarily structurally irrational. However, once you become fully aware that you have attitude *A*, you are certainly disposed to believe that you have attitude *A*. Consequently, Worsnip’s account would incorrectly qualify these combinations as structurally irrational.<sup>11</sup>

Analogously, looking at Lord and Kiesewetter’s proposal, violating a decisive reason turns out to be neither necessary nor sufficient for structural irrationality. Suppose you intend to heal from a deep trauma and you are normatively permitted to intend so. Also, you intend not to smoke and, again, you are permitted to intend so. However, you also believe that you cannot heal from the trauma without smoking. Due to quirky circumstances, suppose you also have sufficient evidence for the truth of this belief and are thus permitted to believe so.

In these circumstances, you are structurally irrational: you fail to intend something you deem necessary for your intended ends. However, your attitudes do not violate a decisive reason. You are permitted (and thus *lack* decisive

- 9 Worsnip, “What Is (In)Coherence?” 198; see also Worsnip, *Fitting Things Together*, sec. 5.4.3.
- 10 Worsnip tries to counter this problem by arguing that the disposition not to sustain structurally irrational patterns of attitudes comes in degrees and that the weaker the disposition, the less irrational the pattern in question. So the fact that “clear-eyed akrasia is possible” just indicates that, in general, akrasia represents a weaker form of structural irrationality: “The most incoherent sets of mental states are ones whereby the disposition is so strong that it cannot be blocked; these sets of states will be impossible to sustain jointly under conditions of full transparency. But in less incoherent cases, such as akrasia, the disposition is weak enough to sometimes be blocked” (Worsnip, “What Is (In)Coherence?,” 200; see also Worsnip, *Fitting Things Together*, sec. 5.4.3).
- 11 For a more complete version of this criticism see Fink, “What (In)Coherence Is Not.” See also Worsnip, *Fitting Things Together*, sec. 5.4.4.

reason not) to have each attitude. Consequently, the violation of a decisive reason does not qualify as a necessary condition for structural irrationality.<sup>12</sup>

These shortcomings reveal that we still lack an account of the essence of structural irrationality. This paper attempts to redeem this situation. I shall offer an original, reductive, and unified account of structural irrationality. The core of the account can be stated as follows: a set of attitudes is structurally irrational if and only if it is metaphysically impossible for those attitudes to be jointly successful. I will show that this account can fully explain the irrationality of some of the paradigmatic instances of structural irrationality.

While it is original, it is important to notice that my proposal also incorporates key aspects of the two accounts discussed above. Lord and Kiesewetter claim that a set of attitudes is structurally irrational if and only if it is impossible for those attitudes to be jointly substantially rational. This is similar to the view I offer in this paper, except that my account picks out structural irrationality not in terms of the impossibility of joint substantial rationality but rather in terms of the impossibility of joint attitudinal success. Also, my account (and its understanding of attitudinal success) incorporates a key element of Worsnip's account. I shall identify necessary success conditions for an attitude via that attitude's constitutive dispositions. That is, *s* is a necessary success condition for an attitude if and only if *A* constitutively aims at *s*.<sup>13</sup>

This paper is organized as follows. Sections 1–4 establish the core conditions of the account I wish to defend. I demonstrate how the idea that the attitudes in an irrational set of attitudes cannot possibly be jointly successful can unify an explanation of the irrationality of contradictory beliefs and intentions, cyclical preferences, and akratic and instrumental incoherence.<sup>14</sup> In section 5, I then add three subjective conditions to the account. Roughly, my proposal will be as follows: a set of attitudes is irrational if and only if it is transparent to the person who has them that it is impossible for them to be jointly successful.

12 In addition, as I have argued elsewhere, the violation of a decisive reason does not qualify as a *sufficient* condition for structural irrationality (Fink, "Structural Irrationality Does Not Consist in Having Attitudes You Ought Not to Have"). Lord and Kiesewetter's reasons-violation view requires a number of assumptions that imply that you violate a decisive reason whenever you intend something you take to be normatively optional. However, intending to *p* and believing that it is neither the case that you ought to *p* nor that you ought to not-*p* does not qualify as a structurally irrational pattern of attitudes. The shortcomings of the two prevalent approaches make it necessary to seek an alternative account of the essence of structural irrationality.

13 I thank an anonymous reviewer for highlighting these similarities.

14 For a first and underdeveloped version of the account developed here, see Fink, "A Constitutive Account of 'Rationality Requires,'" sec. 8.

## 1. A SIMPLE ACCOUNT

This paper aims to establish the essence of structural irrationality. I seek to determine what it is to be structurally irrational and thus what unifies the specific domain of structurally irrational attitudes.

To begin, I will stipulate a simple view of what makes a set of attitudes structurally irrational. This view locates structural irrationality exclusively in the impossibility of the collective success of the attitudes in question. I will refer to this view as the Simple Impossible Success Account (SISA). I will explain what I mean by “success” in section 2, and I will turn to specific examples of irrational sets of attitudes in section 3.

Informally, SISA can be defined as follows:

*SISA Informal:* A set of attitudes is irrational if and only if it is impossible for that set to be successful.<sup>15</sup>

Accordingly, irrationality consists in the impossibility of concurrent attitudinal success. Put differently, it is the necessity of attitudinal failure that makes that set of attitudes irrational. As we will see, SISA will turn out to be overly restrictive and inclusive at the same time. It both identifies as irrational sets of attitudes that are *not* irrational and fails to account for the irrationality of attitudes that *are* irrational. Nevertheless, it is worth gaining a detailed understanding of this account as it brings to light an essential aspect of structural irrationality.

In what follows, I will use “A set of attitudes is irrational” as shorthand for “Necessarily, if you adopt this set of attitudes, then you are not fully rational.” By “a set of attitudes” I mean any possible set of a person’s present and/or absent attitudes. By “present attitude” I mean a mental relationship between a person and a particular object. By “absent attitude” I mean the lack of a particular mental relationship between a person and a particular object. I will use “attitudes\*” to refer to both present and absent attitudes.<sup>16</sup> Unless specified to the contrary, I will also assume that all attitudes\* are contemporaneous.

In order to represent attitudes\*, I will resort to the following schemas:

<A:  $x$ > and <not-A:  $x$ >,

where  $A$  stands for the *type* of attitude (e.g., belief, intention, preference, fear, admiration, hope),  $x$  stands for the *object* of the attitude (e.g., a proposition,

15 To save space, I use “set” here as shorthand for the members of that set taken together (not the abstract entity of the set itself). Thus, “it is impossible for the set to be successful” is meant to express that “it is impossible for the attitudes in that set to be jointly successful.”

16 When I speak of “attitudes,” I will only refer to present attitudes, i.e., attitudes I assume one has. When I speak of “absent attitudes,” I will refer to attitudes one does not have.

state of affairs, person), and “not” signifies the absence of an attitude. For reasons that will become apparent later, I will assume that “ $\langle A: x \rangle$ ” and “ $\langle \text{not-}A: x \rangle$ ” stand not for attitudes\* themselves but for the *propositions* that represent those attitudes\*.

Accordingly, “ $\langle \text{belief: } the\ cat\ is\ on\ the\ mat \rangle$ ” stands for the proposition “You believe that the cat is on the mat.” Likewise, “ $\langle \text{not-intention: } you\ go\ to\ school \rangle$ ” stands for the proposition “It is not the case that you intend to go to school.”

SISA Informal refers to the impossibility of the *success* of a set of attitudes. I will stipulate a general and a substantive view of attitudinal success in section 2. Here, I can say how I will model attitudinal success formally.

I will assume that every attitude comes with a particular set of success conditions. Success conditions relate to attitudes as follows. An attitude picked out by

$\langle A: x \rangle$

is successful only if all success conditions obtain. Suppose, for example, that the propositions

$s_1, s_2 \dots s_n$

denote the success conditions of the attitude picked out by

$\langle A: x \rangle$ .

Then, the success of the attitude represented by  $\langle A: x \rangle$  requires the truth of

$s_1, s_2 \dots s_n$

which I shall refer to as “success propositions.”

I need to add one further clarification. SISA Informal says that a set of attitudes\* is irrational if and only if it is impossible for its members to be jointly successful. I interpret this as follows. First, by “impossible” I mean “metaphysically impossible.” Second, I assume that it is metaphysically impossible for a set of attitudes to be jointly successful if and only if the success propositions of the attitudes in question cannot be true without a contradiction’s being true.<sup>17</sup>

17 By saying that the success propositions of the attitudes in question cannot be true without a contradiction’s being true, I do not mean to say that all instances of metaphysical impossibility *formally* entail a contradiction. For example, (1) “Peter is a bachelor” can be true only if (2) Peter is unmarried. However, 1 does not formally entail 2. (It would do so, of course, if we were to add “If Peter is a bachelor, then Peter is unmarried” to 1 and 2.) This distinction between requiring the truth of a contradiction and formally entailing a contradiction will become important in section 7, when I add transparency conditions to the developed account of structural irrationality.

For example, suppose the success of your attitudes requires the truth of (1) “The numerical value  $x$  is greater than  $y$ ” and (2) “The numerical value  $y$  is greater than  $x$ .” There is no metaphysically possible world in which 1 and 2 are jointly true. In fact, the joint truth of 1 and 2 would require the truth of the flat contradiction that  $x$  is greater than  $y$  and it is not the case that  $x$  is greater than  $y$ .

With these preliminaries in hand, I can now render SISA more precisely. Let  $M$  be a set of attitude\* propositions  $\langle A: x \rangle$  and  $\langle \text{not-}A: x \rangle$ . Then SISA purports to determine the irrationality of the attitudes\* that  $M$  represents as follows. First, assign all success propositions to every individual present attitude  $\langle A: x \rangle$  that  $M$  picks out. Second, form the complete set of success propositions for all attitudes  $\langle A: x \rangle$  that  $M$  picks out. Call this set “ $S_M$ .” Third, examine whether it is metaphysically impossible for all propositions in  $S_M$  to be jointly true. If it is metaphysically impossible, then  $M$  represents an irrational set of attitudes. If it is not metaphysically impossible, then  $M$  does not represent an irrational set of attitudes.

SISA can be formally stated as follows:

*SISA Formal:*  $M$  represents an irrational set of attitudes if and only if  $S_M$  entails<sub>me</sub> a contradiction.

I use “entails<sub>me</sub>” here as a technical term, meaning that it is metaphysically impossible for all propositions in  $S_M$  to be true without a contradiction’s being true. I will contrast “entails<sub>me</sub>” with logical entailment or consequence later. This will become significant when it comes to introducing the kind of transparency required by my account of irrational attitudes.

I argue that SISA can explain and unify a core segment of structural irrationality. In particular, it manages to account for the irrationality of contradictory beliefs and intentions and, as I will show below, some forms of instrumental incoherence as well. Before I can demonstrate this in detail, however, I need to say more about attitudinal success. In particular, I need to state how we can correctly assign particular success propositions to particular attitudes.

## 2. ATTITUDINAL SUCCESS

Attitudes come with success conditions. This is a key assumption of my paper. I argue that understanding attitudinal success is essential to understanding structural irrationality. I operate with an essential or constitutive notion of success here. You may deem your intention to go to a bar a success if doing so results in your meeting the love of your life, but this is a non-constitutive kind of success. Intending to go to a bar and *not* meeting the love of your life (however regrettable) does not necessarily indicate that your intention was essentially defective.

I will distinguish between two types of success propositions. Both are equally relevant to establishing when a set of attitudes is irrational. One type of success proposition will pick out the success of an attitude *qua* its *correctness*. The other type picks out the success of an attitude *qua* its *executive performance*. I will explain below how these two types differ and how I arrive at them.

Before I justify my particular assignment of success propositions, I will first simply list a set of success propositions for the types of attitudes that are involved in the paradigmatic cases of structural irrationality, such as contradictory beliefs and intentions, instrumental and akratic incoherence, and cyclical preferences. Later, when explaining the irrationality of the paradigmatic cases of structural irrationality, I will rely exclusively on the success propositions in **bold**. As I will emphasize below, each listed success condition only states a necessary condition for the full success of the respective attitude. I am committed neither to conceiving of these conditions as sufficient nor to conceiving of them as contributory conditions for attitudinal success.

Let us start with belief. I will assume that a belief that *p* is fully successful only if

***p***.

An intention that *p* is fully successful only if

it is not the case that you ought to not-*p*, and

***p***.

An all-things-considered ought judgment (that expresses a truth-apt cognitive attitude such as a belief) is successful only if

**you ought to *p***, and

*p* is possible, and

**if you ought to *p*, then *p***.

By contrast, an all-things-considered ought judgment (that expresses a non-truth-apt, noncognitive attitude such as a desire or an intention) is successful only if

*p* is possible, and

***p***.

If a preference is sensitive to a comparative judgment (i.e., having the preference depends on a judgment that *a* has more of a certain property, say *F*, than *b*), then the preference is successful only if



***a* is Fer than *b*.**

If a preference subsists in what I shall call a “conditional intention” (i.e., roughly, one intends to [*a* and not-*b*] whenever one will either *a* or *b*), then the preference is successful only if the following material conditional holds true:

**if (*a* or *b*), then (*a* and not-*b*).**<sup>18</sup>

What justifies this assignment of success propositions? I assume that many philosophers will share the intuition behind many of the success propositions specified above. That a false belief is not entirely successful is indeed uncontroversial. That an unrealized intention is not entirely successful is equally uncontroversial. However, some of the other conditions I have specified are in need of explanation. That the material conditional “If (*a* or *b*), then (*a* and not-*b*)” is a success proposition for a particular type of preference may not be immediately obvious.

I offer a method for identifying success propositions. This method is built upon the following principle:

*Success*: Necessarily, *p* is a success proposition of  $\langle A: x \rangle$  if and only if  $\langle A: x \rangle$  constitutively aims at *p*.<sup>19</sup>

I follow Paul Katsafanas’s definition of constitutive aims:

Let *A* be a type of attitude or event. Let *G* be a goal. *A* constitutively aims at *G* iff (i) each token of *A* aims at *G*, and (ii) aiming at *G* is part of what constitutes an attitude or event as a token of *A*.<sup>20</sup>

In short, if an attitude of type *A* aims constitutively at *G*, then part of what makes that attitude a member of type *A* is that it is directed at *G*.

I suggest that *p* is a success proposition of  $\langle A: x \rangle$  if and only if  $\langle A: x \rangle$  constitutively aims or is directed at *p*. In general, there are two distinct ways in which an attitude can aim at something. First, an attitude can aim at *x* by instantiating a disposition to bring about *x*. This is the sense in which, for example,

18 In order to limit the number of variables used in this paper, I use “*p*” (as well as “*a*” and “*b*”) for a range of propositions *and* action types. I hope (and am confident) that the reader will be able to discern this dual use here. What is most important is that whenever “*p*,” “*a*,” and “*b*” are embedded in a complex syntactic construction (for example: “*a* is Fer than *b*”), the resulting clause represents a proposition.

19 I adopt Paul Katsafanas’s suggestion that constitutive aims set up a fundamental and intrinsic standard of attitudinal success. That is: “If *X* [constitutively] aims at *G*, then *G* is a [fundamental or intrinsic] standard of success for *X*” (Katsafanas, *Agency and the Foundations of Ethics*, 39).

20 Katsafanas, *Agency and the Foundations of Ethics*, 39.

an intention that  $p$  aims at its implementation, i.e.,  $p$ . Second, an attitude can aim at  $p$  by having a propensity to be abandoned in the face of not- $p$ . This is the sense in which, for example, a belief that  $p$  aims at truth, i.e.,  $p$ . Of course, for these dispositions to ground a *constitutive* aim, they need to be essential to the attitude in question.

Here is a more precise characterization of these two dispositions. I assume that an attitude  $\langle A: x \rangle$  aims constitutively at  $p$  if and only if either

1.  $\langle A: x \rangle$  disposes\* you to  $p$ , or
2. awareness of not- $p$  disposes\* you to  $\langle \text{not-}A: x \rangle$ ,

where “disposes\*” signifies a disposition that is *essential* to having  $\langle A: x \rangle$ .

The two dispositions differ in a directional sense. The first is an “attitude-to-world” disposition. Here the attitude constitutes a stimulus condition of the disposition. This type of disposition picks out the *executive aim* of the attitude. Failing to meet this aim implies an executive defect on the part of the attitude.

The second disposition is a “world-to-attitude” disposition. Here, the disposition manifests itself in response to (becoming aware of) how the world is. An aspect of the world (or awareness thereof) constitutes a stimulus condition of altering the mind. This type of disposition picks out a condition of *correctness*. Failing to meet this condition implies that the attitude is defective *qua* being incorrect.

Before I turn to justifying the success conditions specified above, let me add a crucial clarification. In order to make sense of these success conditions, it is essential to understand them as merely stating *necessary* conditions for an attitude’s success. I only claim that an attitude is not entirely successful if one of its success conditions turns out to be false. However, I do not wish to make the additional claim that the truth of a success condition is sufficient for or necessarily contributes to the degree of an attitude’s success.

Of course, this will likely be the case for certain success conditions. I would think, for example, that truth contributes to the success of a belief. But in other cases, this will not be so. For example, I identified the material conditional “If you ought to  $p$ , then  $p$ ” as a success condition for a judgment that you ought to  $p$  (which I read as a material conditional). I understand this as saying that if that conditional turns out to be false (i.e., you ought to  $p$ , but you do not  $p$ ), your ought judgment will not be (fully) successful. This strikes me as evident: you judge that you ought to  $p$ , you (in fact) ought to  $p$ , yet  $p$  is never realized. However, I do not mean to say that the truth of this conditional will necessarily contribute to the success of your judgment. This is obviously the case if the condition turns out true just in virtue of its antecedent’s being false (i.e., it is not the case that you ought to  $p$ ).

With this understanding of success propositions in hand, I will now justify the success propositions I assigned above. Let us first look at beliefs. Beliefs are, constitutively, truth-taking attitudes. If you believe that the cat is on the mat yet you become aware that it is not the case that the cat is on the mat, this will dispose\* you to give up your belief that the cat is on the mat. The proposition “The cat is on the mat” thus qualifies as a success proposition of your belief. Or, in general, a belief that  $p$  is successful only if  $p$ .

I now turn to intentions. An intention is “a description of some future action, addressed to the prospective agent, and cast in a form whose point in the language is to make the person do what is described.”<sup>21</sup> Intentions are, constitutively, truth-making attitudes; they aim at implementation. If you intend  $p$ , you are disposed\* to make  $p$  true. An intention that  $p$  is therefore an executive success only if  $p$ ; an intention that  $p$  is successful only if  $p$ .<sup>22</sup>

I now turn to all-things-considered ought judgments. Here, we face an initial difficulty. There are two principal views as to the type of attitude an ought judgment represents. An ought judgment can express a cognitive (and truth-apt) attitude, such as a belief, or it can express a noncognitive (and thus non-truth-apt) attitude, such as an intention or desire. If ought judgments are beliefs, then “You ought to  $p$ ” is a success proposition of your judgment that you ought to  $p$ . This simply follows from the success conditions of ordinary nonnormative beliefs.

Moreover, it is plausible that ought beliefs also have executive success conditions. I assume that if you believe you ought to  $p$ , then you are disposed\* to  $p$ . That is, like intentions, ought beliefs aim at implementation. In this case,  $p$  turns out to be a success proposition of a belief that you ought to  $p$ .

I will be slightly cautious with this condition, however. There may be circumstances where the executive disposition\* of an ought belief does not give rise to this success proposition. Suppose you believe that you ought to  $p$ , yet you become aware that it is not the case that you ought to  $p$ . In this scenario, your awareness may cancel your disposition to  $p$  (in the very least, I am not in a position to exclude this). I will therefore offer a weaker proposal. Your ought

21 Anscombe, *Intention*, 3.

22 Moreover, I assume that “It is not the case that you ought to not- $p$ ” (i.e., the absence of an ought to the contrary) is a success proposition of an intention that  $p$  (although I will not rely on this assumption in explaining paradigmatic cases of structural irrationality). Nishi Shah supports this idea as follows: “My hypothesis is that the concept of intention includes a standard of correctness. Just as classifying an attitude as a belief entails applying to it the standard of being correct if and only if its content is true, likewise classifying an attitude as an intention entails applying to it the standard of being correct if and only if it is not the case that one ought not to perform the action that is its object” (Shah, “How Action Governs Intention,” 12).

belief succeeds only if the material conditional “If you ought to  $p$ , then  $p$ ” is true. That is, a necessary condition of the success of your ought belief is that if you ought to  $p$ , then  $p$ . So, from an executive point of view, your belief that you ought to  $p$  succeeds only if  $p$  or if it is not the case that you ought to  $p$ .

This weakening makes sense. The executive disposition to bring about  $p$  if you believe that you ought to  $p$  may be impaired if you are aware that it is not the case that you ought to  $p$ . I assume that awareness is factive. That is, if you are aware that it is not the case that you ought to  $p$ , then it is not the case that you ought to  $p$ . This is how, under such awareness, a belief that you ought to  $p$  can be successful even if you do not bring about  $p$ .

I now turn to a noncognitivist interpretation of ought judgments. On this interpretation, ought judgments are not truth apt. So, neither “You ought to  $p$ ” nor “If you ought to  $p$ , then  $p$ ” can qualify as a success proposition of your judgment that you ought to  $p$ . Nevertheless, for the noncognitivist, a key constitutive role of an ought judgment is to make you do what you judge you ought to do.<sup>23</sup> Like intentions, ought judgments are motivating states. They are truth-making attitudes: if you judge that you ought to  $p$ , you are disposed\* to bring it about that  $p$ ;  $p$  is thus a success proposition for the judgment that you ought to  $p$ . In fact, that ought judgments aim at implementation seems to be a corollary of the noncognitivist take on “ought.”

I now turn to success with regard to preferences. Preferences, I take it, are dispositions to choose. So, as a minimal aspect of preferences, you prefer  $a$  over  $b$  only if you are in a mental state that is prone to cause you to  $a$  if you are prompted to choose between  $a$  and  $b$ .<sup>24</sup> On this view, there is an inherent difficulty in assigning success propositions to preferences. First, non-attitudinal mental states can constitute preferences (for example, a non-attitudinal perceptual state may dispose you to choose  $a$  over  $b$ ). Second, there are a variety of attitudes that can constitute a disposition that amounts to a preference.

Suppose you prefer cycling to driving. Your preference may consist in a (comparative) desire. That is, you like cycling more than driving. Or it may be an intention-belief pair. You may intend to live healthily, and you believe that cycling is healthier than driving. Or this may be due to a comparative judgment, e.g., that cycling is healthier, better for the environment, or in any other comparative sense *Fer* than  $b$ . Or it may be a directly comparative evaluative belief. You may believe that cycling is better than driving. Alternatively, it may be what I call

23 Typically, noncognitivists treat this property of ought judgments as key evidence for their noncognitive status.

24 Cf., e.g., Rabinowicz, “Value Relations” and “Modeling Parity and Incomparability.”

a “conditional intention”: you intend to cycle if you will either cycle or drive. I assume that all such attitudes can constitute a preference for cycling over driving.

David Hume declared that “a passion must be accompany’d with some false judgment, in order to its being unreasonable; and even then ’tis not the passion, properly speaking, which is unreasonable, but the judgment.”<sup>25</sup> In a similar vein, I suggest that the success of a preference derives from the attitude that constitutes it. Thus, I do not think it necessary to list success propositions of preferences that are based on (normative) beliefs and/or (ordinary) intentions. If a belief and/or an intention amount(s) to a preference, then the success proposition of the resulting preference is simply the success proposition of the belief and/or the intention. For example, if a belief that *a* is healthier than *b* partly constitutes a preference for *a* over *b*, then one success proposition of this preference is the success proposition of the corresponding belief.

Thus far, I have refrained from ascribing success conditions to desires. This is a tricky issue. Extreme Humeans believe “that rationality allows a person to have any pattern of preferences whatsoever.”<sup>26</sup> If this is correct, then this severely restricts the possibility of assigning success conditions to desires. If irrationality consists in the impossibility of success, then desires do not come with success conditions that could ever be necessarily incompatible. I will not dive into the extreme Humean abyss, however. Instead, I will put forward a moderate proposal. Take a preference for *a* over *b* that consists in a comparative desire. That is, you desire *a* more than *b*. I will assume that your preference comes with a disposition\* to discard the preference if you become aware that not-*p*. So, analogously to how I identify success propositions above, *p* is a success proposition for that type of preference.

I will deliberately refrain from identifying any particular property here. I am only claiming that there is an evaluative and comparative property for which the discovery that *a* does not have more of that property than *b* would necessarily incline you to abandon your preference for *a* over *b*. For many philosophers, this suggestion will be more than natural. Anscombe suggested, for example, that desires aim at the good: “The conceptual connexion between ‘wanting’ ... and ‘good’ can be compared to the conceptual connexion between ‘judgment’ and ‘truth.’ Truth is the object of judgment, and good the object of wanting.”<sup>27</sup>

With these three assumptions in hand, it is only a small step to construe a success proposition for comparative desires. Suppose you prefer *a* to *b* based

25 Hume, *A Treatise of Human Nature*, III.iii.3.

26 Broome, *Ethics out of Economics*, 76.

27 Anscombe, *Intention*, 76.

on a comparative desire. If the conceptual connection between “wanting” and “good” is analogous to the conceptual connection between “judgment” and “truth,” it is natural to suppose that you are disposed\* to give up your preference for *a* over *b* should you become aware that it is not the case that *a* is better than *b*. In this case, “*a* is better than *b*” turns out to be a success proposition of your preference for *a* over *b*. (Of course, this suggestion only works if the betterness relation holds independently of your preference; that is, “*a* is better than *b*” does not imply that you prefer *a* to *b*.)

In this paper, however, I need not rely on the assumption that desires aim at the good or that comparative desires aim at the better (although I am not denying this either). My proposal is slightly more modest and general. There only needs to be *some* relation “\_\_\_ is *F*er than \_\_\_” that holds between *a* and *b* for a comparative desire preference to succeed. That relation must be such that the realization of its absence necessarily disposes\* you to drop the comparative desire. *F* may very well be some kind of goodness (desirability or choiceworthiness), but I will not commit myself to this assumption.

What about preferences that consist in a “conditional intention”? Following Ralph Wedgwood’s suggestion, a conditional intention is a (complex) intention to do one thing—say *a*—and not another—say *b*—if you will do either *a* or *b*.<sup>28</sup> (I understand the “or” here as an exclusive disjunction, meaning that the choice between *a* and *b* is an exclusive one.) When going to a restaurant, for example, you may intend to eat monkfish and not chicken if you narrow down your choice to monkfish or chicken. Or when buying a used car, you may intend to buy a Saab 9-3 and not a Saab 9-5 if given a choice between the two. In this case, you have a conditional intention to buy a Saab 9-3. This intention exemplifies a preference: you prefer buying a Saab 9-3 to buying a Saab 9-5.

When do such preferences succeed? Suppose you prefer *a* to *b* because of a conditional intention to *a* given that *a* and *b* are your only options. With one exception, we cannot suggest (as I have done above for intentions *simpliciter*) that this intention is successful only if *a* is the case.<sup>29</sup> As a vegetarian, for example, you generally prefer eating no beef to eating beef. Yet when faced with the exclusive choice between beef and nothing (i.e., you will eat beef or nothing),

28 “To form a preference for a proposition over the relevant alternatives is not necessarily to form a choice or intention to realize that proposition; it would at most be to form a *conditional* intention—in effect, the intention of acting in such a way that the proposition in question is true, rather than in such a way that the relevant alternative is true, *if one does either*” (Wedgwood, *The Nature of Normativity*, 120).

29 Here is the exception: suppose you prefer *a* to not-*a*. Then the success proposition will be “If *a* or not-*a*, then *a* and not-not-*a*,” which is equivalent to *a*. This is relevant to some of the examples I present below.

you prefer to eat beef to eating nothing. That does not mean, of course, that your preference succeeds *only* if you end up eating beef (i.e., eating beef is not a necessary condition for the success of this preference).

I propose “If ( $a$  or  $b$ ), then ( $a$  and not- $b$ )” as a success proposition for a preference ( $a$  over  $b$ ) that is based on a conditional intention. This is a natural suggestion. A constitutive aim of such a preference is to bring about ( $a$  and not- $b$ ) whenever you either  $a$  or  $b$ . This is only the case if the following material conditional holds true: “If ( $a$  or  $b$ ), then ( $a$  and not- $b$ )” (where “or” and “and” signify the standard truth-functional (inclusive) disjunction and conjunction).

Here is the same point from another angle. When is a conditional intention (to  $a$  and not- $b$  if you will do either  $a$  or  $b$ ) *not* successful? The answer seems obvious. It is unsuccessful if  $b$  is true while  $a$  is not true. Suppose, again, that you have a conditional intention to eat beef if you will either eat beef or eat nothing, yet you end up eating nothing. Then your conditional intention will not be successful. That is why, when it comes to a preference for  $a$  over  $b$  that is based on a conditional intention, in order for it to succeed, “If ( $a$  or  $b$ ), then ( $a$  and not- $b$ )” must hold true (which requires that either both  $a$  and  $b$  are false or ( $a$  and not- $b$ ) is true).

It is fortunate, though, that we do not need to handle such a cumbersome success proposition. This is because “If ( $a$  or  $b$ ), then ( $a$  and not- $b$ )” turns out to be equivalent to “not- $b$ .” It is easy to show this. The truth of  $b$  suffices for the falsity of “If ( $a$  or  $b$ ), then ( $a$  and not- $b$ ),” for it guarantees the truth of its antecedent and the falsity of the consequent. Likewise, the falsity of  $b$  suffices for the truth of “If ( $a$  or  $b$ ), then ( $a$  and not- $b$ ).” If both  $b$  and  $a$  are false, then the antecedent is false and hence the conditional is true. If  $b$  is false and  $a$  is true, then both the antecedent and the consequent are true. So, the conditional is again true. Consequently, the negation of  $b$ —not- $b$ —turns out to be logically equivalent to “If ( $a$  or  $b$ ), then ( $a$  and not- $b$ ).” Thus, I will treat not- $b$  as a success proposition for a preference for  $a$  over  $b$  that is based on a conditional intention to  $a$  and not- $b$  if you will do either  $a$  or  $b$ .

To summarize this section, I have created a table of the success conditions proposed above (table 1). In this table,  $O$  refers to “You ought to . . .”;  $\diamond$  refers to “It is possible that”; and *Fer* signifies a comparative relation. “—” signifies that I will refrain from specifying a particular success proposition (this is not to say, of course, that there is not one). “. . .” is meant to indicate that there could be more success propositions than those I have specified. Again, in explaining the irrationality of the paradigmatic cases of structural irrationality, I will rely exclusively on the success propositions in **bold**. I have listed the others for illustrative purposes only.

Table 1

<Attitude Type: Object>	Correctness Success	Executive Success
<belief: $p$ >	$p, \dots$	—
<intention: $p$ >	not- $O$ (not- $p$ ), ...	$p, \dots$
<ought judgment: $p$ ><cognitive>	$Op, \diamond p, \dots$	if $Op$ , then $p, \dots$
<ought judgment: $p$ ><noncognitive>	$\diamond p, \dots$	$p, \dots$
<preference: $a$ over $b$ ><comparative desire>	<b><math>a</math> is Fer than <math>b, \dots</math></b>	not- $b, \dots$
<preference: $a$ over $b$ ><conditional intention>	—	<b>not-<math>b, \dots</math></b>

3. APPLYING THE SIMPLE ACCOUNT

Let us return to SISA. SISA says that a collection of attitudes\* is irrational if and only if their success propositions entail<sub>me</sub> a contradiction. Recall that “entail<sub>me</sub> a contradiction” means that it is metaphysically impossible for a set of propositions to be true without a contradiction’s also being true. In this section, I will demonstrate that this account can unify an explanation of the irrationality of contradictory beliefs, contradictory intentions, and cyclical preferences. In the next section, I will show how two modest modifications of SISA render it able to explain the irrationality of akratic incoherence and two types of instrumental incoherence.

Let us begin with contradictory beliefs:

$$M_1 \{ \langle \text{belief: } p \rangle; \langle \text{belief: not-}p \rangle \}.$$

Let us form  $M_1$ ’s set of success propositions. I assume that  $\langle \text{belief: } p \rangle$  is successful only if  $p$ . Correspondingly,  $\langle \text{belief: not-}p \rangle$  is successful only if not- $p$ .  $M_1$ ’s set of success propositions,  $S_{M_1}$ , reads as follows:

$$\{ p; \text{not-}p \}.$$

This set most evidently entails<sub>me</sub> a contradiction. The success of the belief that  $p$  thus precludes the success of the belief that not- $p$  (and *vice versa*). According to SISA,  $M_1$  is thus irrational.

Next is  $M_2$ :

$$M_2 \{ \langle \text{intention: } p \rangle; \langle \text{intention: not-}p \rangle \}.$$

Let us consider  $M_2$ ’s set of success propositions. I assume that  $\langle \text{intention: } p \rangle$  is successful only if  $p$ . Correspondingly,  $\langle \text{intention: not-}p \rangle$  is successful only if not- $p$ .  $S_{M_2}$  thus reads as follows:

$$\{ p; \text{not-}p \}.$$



$S_{M_1}$  and  $S_{M_2}$  are identical. *SISA* also implies that  $M_2$  is irrational. The success of an intention that  $p$  precludes the success of an intention that not- $p$  (and *vice versa*).

I now turn to cyclical preferences. Recall  $M_3$ :

$M_3$  { <preference:  $a$  to  $b$ >; <preference:  $b$  to  $c$ >; <preference:  $c$  to  $a$ > }.

I said in section 2 that the success of a preference derives from the attitudes that constitute it. Suppose, first, that the preferences in  $M_3$  are based on comparative desires. That is, you desire  $a$  more than  $b$ ,  $b$  more than  $c$ , and  $c$  more than  $a$ . I assume that a comparative desire for  $a$  over  $b$  is successful only if  $a$  is *Fer* than  $b$ , where *F* stands for a comparative property (i.e., anything  $a$  and  $b$  can have more or less of).

This generates the following set of success propositions  $S_{M_3}$ :

{ $a$  is *Fer* than  $b$ ;  $b$  is *Fer* than  $c$ ;  $c$  is *Fer* than  $a$ }.

These propositions entail<sub>me</sub> a contradiction. The “\_\_\_ is *Fer* than \_\_\_” relation is necessarily acyclical.<sup>30</sup> The truth of any two of these propositions precludes the truth of the third. A set of cyclical comparative desire-based preferences is such that it is impossible for the set to succeed. In this sense, *SISA* can already explain the irrationality of cyclical preferences.

I now turn to preferences constituted in conditional intentions. Here, things are slightly more complicated. Suppose that  $M_3$ 's preferences are constituted by conditional intentions. I assume that a conditional intention to  $a$  and not- $b$  if either  $a$  or  $b$  is successful only if not- $b$  is true. Accordingly, for  $M_3$ , this generates the following set of success propositions  $S_{M_3}$ :

{not- $b$ ; not- $c$ ; not- $a$ }.

Does this set entail<sub>me</sub> a contradiction? In fact, this hinges on one condition. Suppose that, as a matter of metaphysical necessity, at least one of  $a$ ,  $b$ , or  $c$  is an exhaustive proposition. That is, at least one of them is necessarily true. Consequently, the propositions in  $S_{M_3}$  could never be true together; their joint truth would require the truth of a contradiction, and so  $S_{M_3}$  would entail<sub>me</sub> a contradiction.

Consequently, *SISA* implies that a set of cyclical conditional intentions is structurally irrational whenever we consider mutually contrary propositions. However, if it is possible for *none* of the options to be realized, then *SISA* does not predict structural irrationality. I admit that this is a considerable drawback for *SISA*. If conditional intentions are genuine preferences, then there seem to

30 Cf. Broome, *Weighing Lives*, 51. For a famous disagreement, see Tempkin, “Intransitivity and the Mere Addition Paradox.”

be sets of cyclical preferences that do not turn out to be structurally irrational. This is so for sets of cyclical preferences where it is possible not to realize any of the options involved.

However, I believe there are two somewhat satisfactory ways to fix this problem. Suppose you have a set of cyclical preferences. You prefer  $a$  to  $b$ ,  $b$  to  $c$ , and yet  $c$  to  $a$ . Suppose your preferences consist of conditional intentions. In addition, let us assume one of two things: either you hold a meta-preference for *either* over *neither* of these options (i.e., you prefer  $[a \text{ or } b \text{ or } c]$  to  $\text{not-}[a \text{ or } b \text{ or } c]$ ) or you believe that  $(a \text{ or } b \text{ or } c)$  will be the case.

Let us look at the meta-preference first. I will assume that it is a conditional intention, too. So, you intend to  $(a \text{ or } b \text{ or } c)$  and *not not-* $(a \text{ or } b \text{ or } c)$  (double negation) if you will either  $(a \text{ or } b \text{ or } c)$  or  $\text{not-}(a \text{ or } b \text{ or } c)$ . Since  $[(a \text{ or } b \text{ or } c) \text{ or } \text{not-}(a \text{ or } b \text{ or } c)]$  is a tautology and *not not-* $(a \text{ or } b \text{ or } c)$  is simply  $(a \text{ or } b \text{ or } c)$ , this meta-preference comes down to a straightforward intention to  $(a \text{ or } b \text{ or } c)$ . The success proposition for such an intention is  $(a \text{ or } b \text{ or } c)$ . Second, suppose you believe that either  $a$  or  $b$  or  $c$  will be the case. The success proposition for such a belief is also  $(a \text{ or } b \text{ or } c)$ .

Let us add this success proposition to  $S_{M_3}$ . This results in:

$\{\text{not-}b; \text{not-}c; \text{not-}a; a \text{ or } b \text{ or } c\}$ .

This set entails<sub>me</sub> a contradiction. If  $\text{not-}b$ ,  $\text{not-}c$ , and  $\text{not-}a$  are true, then  $(a \text{ or } b \text{ or } c)$  is false. Likewise, if  $(a \text{ or } b \text{ or } c)$  is true, then either  $\text{not-}b$  or  $\text{not-}c$  or  $\text{not-}a$  is false. Hence, your preferences cannot succeed jointly. SISA thus explains the irrationality of conditional intention cyclical preferences *if* you also prefer *either* to *neither* of the involved options or you believe that at least one of the options will be realized.

What should we make of this result? By and large, I believe it makes sense. Suppose you have a set of cyclical preferences that are conditional intentions, but you prefer that they not be realized or you believe they will not happen. It seems to me that this would alleviate the irrationality of your intention-based cyclical preferences (at least from an executive point of view). However, if you do prefer any of these options to none of them or believe they will be realized, then you hold a set of attitudes that are structurally irrational.

#### 4. APPLYING AN EXTENDED ACCOUNT

I said that contradictory beliefs, contradictory intentions, cyclical preferences, and instrumental and akratic incoherence are paradigmatic of structural irrationality. Any credible account of structural irrationality needs to explain why these combinations of attitudes are irrational. However, contradictory beliefs

( $M_1$ ), contradictory intentions ( $M_2$ ), and cyclical preferences ( $M_3$ ) demarcate the explanatory limit of SISA. I have shown that, by and large, SISA can explain and unify the irrationality of  $M_1$ ,  $M_2$ , and  $M_3$ . With regard to akratic and instrumental incoherence, however, SISA remains inapt. I will therefore introduce and argue for an extension of SISA.<sup>31</sup>

I loosely defined akratic incoherence as a state in which you believe that you ought to do something without intending to do so. More precisely, you are akratically incoherent if and only if you are in a state where (1) you judge that, all things considered, you ought to  $p$ , (2) you believe that  $p$  only if you intend  $p$ , and (3) yet you have no intention to  $p$ . That is:

( $M_4$ ) {<ought judgment:  $p$ >; <belief:  $p$  only if <intention:  $p$ >>; <not-intention:  $p$ >},

where “only if” is meant to cover any necessary condition that suffices to make  $M_4$  structurally irrational. Here is why 2 is an integral part of akratic incoherence. Suppose you judge that you ought to relax. But you also know that intending to relax is not necessary for relaxing. In fact, an intention to relax will make it much less likely that you will relax. In such circumstances, you are not structurally irrational if you believe you ought to  $p$  yet you refrain from intending  $p$ .<sup>32</sup> Thus, to be genuinely akratic, you must affirm that you will not end up doing what you believe you ought to do unless you intend so.<sup>33</sup>

I roughly defined instrumental incoherence as failing to intend something you deem necessary for your intended ends (instrumental incoherence). More precisely, I will assume that one type of instrumental incoherence (I will discuss

31 By this I do not mean that SISA can only explain the irrationality of  $M_1$ ,  $M_2$ , and  $M_3$ . It may also explain the irrationality of incompatible credences, for example (or, more generally, the assignments of incompatible probabilities to propositions). Suppose ( $a$  or  $b$  or  $c$ ) are mutually exclusive and exhaustive. Suppose you assign a particular probability to the truth of these propositions. Then it seems that rationality requires that  $a$ ,  $b$ , and  $c$  must add up to one in your assignment of probabilities. One straightforward way to include this under SISA would be to say that the attitude that constitutes the assignment of probability to a proposition (i.e., the corresponding credence or belief) is successful only if it corresponds to the objective probability that the proposition will be true. If you now, for example, assign a probability of 0.3 to  $a$ ,  $b$ , and  $c$ , then this will imply that at least one of your attitudes will not be fully successful. This holds, in fact, for all probability assignments that do not add up to one.

32 Cf. Broome, *Rationality through Reasoning*, 170–72, and “Enkrasia,” 433–34; Kiesewetter, *The Normativity of Rationality*, 192.

33 I am also implicitly stipulating a restriction on your judgments that you ought to  $p$  (<ought judgment:  $p$ >) to judgments where you also believe that  $p$  is in your power and can be brought about by your intentions and/or actions (cf., e.g., Broome, “Enkrasia,” 433–34).

another type later) consists in (1) intending  $p$ , (2) believing that an intention to  $q$  is necessary for realizing  $p$ , yet (3) not intending  $q$ . That is:

$(M_5) \{ \langle \text{intention: } p \rangle; \langle \text{belief: } p \text{ only if } \langle \text{intention: } q \rangle \rangle; \langle \text{not-intention: } q \rangle \}$ .

Let us consider  $M_4$ 's and  $M_5$ 's sets of success propositions. Let us first assume cognitivism about ought judgments.  $M_4$ 's set of success propositions— $S_{M_4}$ —thus reads as follows:

$\{Op; \text{ if } Op, \text{ then } p; p \text{ only if } \langle \text{intention: } p \rangle \}$ .

If the ought judgment is noncognitive,  $S_{M_4}$  reads:

$\{p; p \text{ only if } \langle \text{intention: } p \rangle \}$ .

Likewise,  $M_5$ 's set of success propositions— $S_{M_5}$ —reads:

$\{p; p \text{ only if } \langle \text{intention: } q \rangle \}$ .

On the face of it, none of these sets entails<sub>me</sub> a contradiction. Unless some propositions within a set happen to be metaphysically incompatible with each other (which is, of course, not necessary), it is metaphysically possible for all propositions to be true (without the truth of a contradiction).

This demonstrates a significant limitation of SISA.  $M_4$  and  $M_5$  are structurally irrational, but for almost all common instances of  $M_4$  and  $M_5$ , SISA cannot explain this. We need to extend SISA if we want to utilize its core idea in unifying an explanation of structural irrationality.

SISA says that  $M$  represents an irrational set of attitudes if and only if  $S_M$  (i.e., the set of success propositions of the attitudes represented by  $M$ ) entails<sub>me</sub> a contradiction. Let me apply a subtle extension of this. I propose that when determining whether it is possible for your attitudes to succeed, we need to consider not only current but also *absent* attitudes. Looking back at  $M_5$ , for example, SISA only considers the success propositions of the attitudes in  $M_5$ , i.e.,  $\{p; p \text{ only if } \langle \text{intention: } q \rangle \}$ . However, it ignores the fact that  $M_5$  also contains an absent attitude, i.e.,  $\langle \text{not-intention: } q \rangle$ . I argue that once we modify SISA to include this absent attitude, it will be able to explain the irrationality of  $M_4$  and  $M_5$  (as well as  $M_1$ ,  $M_2$ , and  $M_3$ ).

Spelled out slightly more formally, I propose that  $M$  represents an irrational set of attitudes\* if and only if the *union* of  $M$  and  $S_M$  entails<sub>me</sub> a contradiction. That is, if by merging the sets  $M$  and  $S_M$  we gain a set for which it is metaphysically impossible that all propositions are true at the same time, then and only then is the set of attitudes\* represented by  $M$  irrational. Let us call this the “extended impossible success account” (hereafter EISA). A formal version reads as follows:

*EISA Formal:*  $M$  represents an irrational set of attitudes\* if and only if  $M \cup S_M$  entails<sub>me</sub> a contradiction,

where  $\cup$  symbolizes the union of two sets. Accordingly, a set of attitudes\* represented by  $M$  is irrational precisely when conjoining the propositions that represent those attitudes\* with the propositions that represent the success of the attitudes in  $M$  entails<sub>me</sub> a contradiction. Informally, EISA can be stated as follows:

*EISA Informal:* A set of attitudes\* is irrational if and only if holding those attitudes\* precludes the joint success of the present attitudes in that set,

where “precludes” refers to metaphysical impossibility. That is, for  $M$  to represent an irrational set of attitudes, it need not be *per se* impossible for the present attitudes represented by  $M$  to succeed jointly. Rather, the idea is that it is impossible both for a person to hold the attitudes\* picked out by  $M$  and for the present attitudes represented by  $M$  to succeed jointly.

Can this account explain the irrationality of  $M_4$  and  $M_5$ ? First, let us form the union of  $M$  and  $S_M$ . We simply need to add <not-intention:  $p$ > to the two sets (“cognitive” and “noncognitive”) of success propositions  $S_{M_4}$  and <not-intention:  $q$ > to the set of success propositions  $S_{M_5}$ . The result reads as follows:

{ $Op$ ; if  $Op$ , then  $p$ ;  $p$  only if <intention:  $p$ >; <not-intention:  $p$ >};  
 { $p$ ;  $p$  only if <intention:  $p$ >; <not-intention:  $p$ >};  
 { $p$ ;  $p$  only if <intention:  $q$ >; <not-intention:  $q$ >}.<sup>34</sup>

All three sets entail<sub>me</sub> a contradiction. Consider the first set:  $Op$  and if  $Op$ , then  $p$  formally entail  $p$ . Conjoining  $p$  with  $p$  only if <intention:  $p$ > formally entails <intention:  $p$ >. Conjoining <intention:  $p$ > with <not-intention:  $p$ > then entails the contradiction.

Here is what this means concretely. Consider *akratic incoherence* ( $M_4$ ). Suppose you judge that you ought to  $p$  and you believe that you can  $p$  only if you intend  $p$ . I assume here that your ought judgment expresses a belief. Suppose that both the normative judgment and the belief are successful. That is,

you ought to  $p$ ;

if you ought to  $p$ , then  $p$ ;

and so  $p$  holds true. Moreover, it also holds true that

34 Strictly speaking, the three sets do not represent the union of  $M$  and  $S_M$  because (to avoid unnecessary complexity) I have not included all present attitudes. The result remains the same, however.

if  $p$ , you intend  $p$ .

However,

it is not the case that you intend  $p$ .

This implies, necessarily, that as long as you hold the combination of attitudes\* captured by  $M_4$ , either your normative judgment (that you ought to  $p$ ) or your belief (that if  $p$ , you intend  $p$ ) will not succeed. This is what makes  $M_4$  irrational. Of course, the same result holds if we employ the noncognitive success proposition for ought judgments (compare the second set above).

The third set also entails a contradiction—namely, <intention:  $q$ > and <not-intention:  $q$ >. Suppose you intend  $p$  and you believe that  $p$  only if you intend  $q$ . Suppose that both the intention and your belief are successful. That is, both  $p$  and ( $p$  only if you intend  $q$ ) hold true. The truth of these two propositions implies that you intend  $q$ . But you do not intend  $q$ . Hence, either your intention or your belief will not succeed. This is what makes *means-end absent intention incoherence* ( $M_5$ ) irrational. In sum, EISA can track the irrationality of attitudinal combinations that contain absent attitudes.<sup>35</sup>

Finally, let us turn to another type of *instrumental incoherence*. Here, it is not an absent attitude that precludes the success of your attitudes, but rather a present attitude. Suppose you intend to go shopping. You believe that you will go shopping only if you do not intend to stay at home, and yet you intend to stay at home. Or more formally:

$(M_6)$  {<intention:  $p$ >; <belief:  $p$  only if <not-intention:  $q$ >>; <intention:  $q$ >}.

I suggest that this pattern of attitudes is structurally irrational. Let us first form the set of success propositions  $S_{M_6}$ :

{ $p$ ;  $p$  only if <not-intention:  $q$ >;  $q$ }.

As with  $M_4$  and  $M_5$ , this set does not entail<sub>me</sub> a contradiction. From  $p$  and  $p$  only if <not-intention:  $q$ > we can derive <not-intention:  $q$ >. Conjoined with  $q$ , however, this does not entail<sub>me</sub> a contradiction. This holds under most replacements of  $p$  and  $q$ . So, ordinarily, your attitudes do not undermine their own success.

35 I am grateful to an anonymous reviewer for pointing out that John Brunero offers a similar explanation of the structural irrationality of means-end absent intention incoherence ( $M_5$ ) (Brunero, *Instrumental Rationality*, 178, 197–98). However, unlike myself, Brunero remains skeptical of the claim that constitutive aims or the success of attitudes can explain the structural irrationality of other combinations of attitudes, such as *akratic incoherence* ( $M_4$ ) or *cyclical preferences* ( $M_3$ ) (*Instrumental Rationality*, 178, sec. 7.2.2).

Unlike  $M_4$  and  $M_5$ ,  $M_6$  does not involve an absent attitude. But it involves another element that EISA considers—namely, *present* attitudes. EISA states that we should form the union of  $M$  and  $S_M$ . This includes combining the attitude propositions of the present attitudes in  $M$  with their success propositions.

$M_6$  contains three present attitudes (two intentions and a belief). For the sake of simplicity, we only need to add one of these to  $S_{M_6}$  to generate a contradiction—namely, the intention to  $q$ . The resulting set reads as follows:

$\{p; p \text{ only if } \langle \text{not-intention: } q \rangle; q; \langle \text{intention: } q \rangle\}$ .

This set entails<sub>me</sub> a contradiction. From

$p$

and

$p \text{ only if } \langle \text{not-intention: } q \rangle$

we can infer  $\langle \text{not-intention: } q \rangle$ . Conjoined with

$\langle \text{intention: } q \rangle$

we arrive at a contradiction. So, according to EISA,  $M_6$  turns out to be irrational.

What generates the irrationality of  $M_6$ ? Again, unlike  $M_1$ – $M_3$ , it is not the *success* of some of your attitudes that defeats the success of some of your other attitudes. Likewise, it is not the absence of an attitude (as with  $M_4$  and  $M_5$ ) that defeats an attitude's success. Instead, with  $M_6$  it is the *presence* of an attitude that undermines the success of another attitude. The mere presence of  $\langle \text{intention: } q \rangle$  makes the joint success of  $\langle \text{intention: } p \rangle$  and  $\langle \text{belief: } p \text{ only if } \langle \text{not-intention: } q \rangle \rangle$  impossible. So, in contrast to  $M_4$  and  $M_5$ , you cannot overcome the irrationality of  $M_6$  by adding another attitude to it. As with  $M_1$ – $M_3$ , you need to eliminate at least one of your attitudes from the arrangement.

Let us take stock of where we are thus far. I have proposed that a set of attitudes\* is irrational precisely when it is metaphysically impossible for the conjunction of the following to be true:

1. The propositions that pick out the present attitudes of the set
2. The propositions that pick out the absent attitudes of the set
3. The propositions whose truth is necessary for the success of the present attitudes of the set

In short, if a set of attitudes\* is structured such that it is metaphysically impossible both to hold those attitudes\* and for the present attitudes in that set to succeed, then (and only then) is that set of attitudes\* irrational.

## 5. STRUCTURAL IRRATIONALITY AND REFLECTIVE ACCESSIBILITY

I have shown that EISA manages to unify the six diverse yet paradigmatic instances of irrational patterns of attitudes\*  $M_1$ – $M_6$ . We now have *one* explanation of the irrationality of contradictory beliefs, contradictory intentions, cyclical preferences, and akratic and (two types of) instrumental incoherence. This is great progress indeed. EISA captures essential knowledge about what it is for a set of attitudes to be irrational.

However, does EISA as it stands state the real essence of structural irrationality? Or does this account require further refinements? Here is one issue on which this will depend.

Suppose you believe that  $2 = 1$  or you intend to be married to a bachelor:

<Belief:  $2 = 1$ >;

<Intention: You are married to a bachelor>.

No doubt, your belief and your intention are substantially flawed. They are criticizable in many ways. You believe (or intend) a metaphysical impossibility. But does either of these attitudes suffice to make you structurally irrational (i.e., you are necessarily rendered structurally irrational if you adopt one of these two attitudes)?

If your answer is yes, then this supports the view that EISA captures the essence of structural irrationality. Consider the success proposition of your belief and your intention, respectively:

< $2 = 1$ >;

<You are married to a bachelor>.

Both propositions entail<sub>me</sub> a contradiction:  $2 = 1$  requires not- $(2 = 1)$ ; “You are married to a bachelor” requires the existence of a person who is married *and* not married (i.e., a bachelor).<sup>36</sup> Consequently, EISA implies that both attitudes are necessarily structurally irrational.

However, I believe there are good reasons not to count these attitudes as necessarily structurally irrational. There are mental environments (strange as they may be) in which such individual attitudes may turn out not to be structurally irrational. Suppose you have adopted a worldview with alternative axioms of arithmetic, or one that allows some contradictions to be true. Or you simply

36  $2 = 1 \mid -1$ ;  
 $1 = 0$   
 not- $(1 = 0)$ ;  
 So: not- $(2 = 1)$ .



lack the conceptual or logical capacity to discern that you believe or intend something impossible. Indeed, I assume there are exceptional circumstances in which these may not be structurally irrational. EISA therefore turns out to be overly inclusive: it identifies as irrational combinations of attitudes that are perfectly rational.

In this final section, I will argue that there is one main reason for this. EISA does not presuppose that one has *reflective access* to the fact that one's attitudes cannot succeed jointly. For a set of attitudes to be structurally irrational, however, a person must, at least to some degree, be able to identify the fact that something about her attitudes is defective. Even if your attitudes cannot succeed, charging you with irrationality remains unwarranted as long as you are either genuinely unable or justified in failing to detect this.

In the remainder of this paper, I will therefore argue for three key refinements to EISA. I will defend three conditions that are necessary for the structural irrationality of a set of your attitudes:

1. The impossibility of joint success is transparent to the degree of a logically valid inference.
2. If you were to deploy your full logical abilities, you would be able to detect that the joint success of the attitudes implies a contradiction.
3. You do not have a justified paraconsistent belief that this contradiction is in fact true.

### 5.1. The Traditional Solution

Before I turn to defending these three conditions, I will briefly examine the traditional answer as to why individual attitudes do not count as structurally irrational. I will also explain why it is not a good idea to make EISA subject to the traditional view.

The traditional hypothesis is that structural irrationality is essentially *relational*. That is, structural irrationality, as the name indicates, can arise only from a mismatch among attitudes\*. In an early contribution to this debate, Tim Scanlon explicates this point as follows. Claims of structural irrationality are

structural because they are claims about the *relations* between an agent's attitudes that must hold insofar as he or she is not irrational, and the kind of irrationality involved is a matter of conflict between these attitudes.<sup>37</sup>

37 Scanlon, "Structural Irrationality," 84–85 (emphasis added).

Or, as Niko Kolodny succinctly puts it: “Subjective rationality is a matter of the relations among one’s attitudes.”<sup>38</sup> More recent characterizations tie in with this: Worsnip specifies structural rationality as having “attitudes that are not jointly incoherent.”<sup>39</sup> Kiesewetter writes that “structural irrationality is a kind of irrationality that we can detect simply by looking at a particular combination of attitudes that a person holds.”<sup>40</sup> In sum, irrationality stems from a particular incoherence, disunity, mismatch, or lack of fit *between* attitudes.<sup>41</sup>

The core point of these characterizations seems to be this. For a set of attitudes to be irrational, that set must at least contain either (1) two *antagonistic* attitudes or (2) one attitude and the absence of a *complementary* attitude. For example, the set {<belief:  $p$ >; <belief: not- $p$ >} serves as a fitting illustration of 1: <belief:  $p$ > and <belief: not- $p$ > represent two antagonistic attitudes. The set {<ought judgment:  $p$ >; <belief:  $\Diamond p \Rightarrow$  <intention:  $p$ >>; <not-intention:  $p$ >} serves as a fitting illustration of 2: the absence of <intention:  $p$ > represents the absence of a complementary attitude.

Should we accept this? First, it would satisfy the desideratum of excluding single attitudes from being irrational. By definition, a single attitude lacks (1) the presence of antagonistic and (2) the absence of complementary attitudes.<sup>42</sup> Second, it would be easy to make EISA sensitive to this assumption. We simply need to add the following condition to the account:

*Multiple:* It is possible to generate the contradiction that  $M \cup S_M$  entails<sub>me</sub> from multiple attitudes\* (i.e., either at least two present or one present and one absent attitude) in  $M$ .<sup>43</sup>

38 Kolodny, “Why Be Rational?,” 530.

39 Worsnip, review of *The Normativity of Rationality*.

40 Kiesewetter, *The Normativity of Rationality*, 14.

41 Reisner, “Is the Enkratic Principle a Requirement of Rationality?” Of course, interpreted strictly, Scanlon’s point is perplexing. Irrationality cannot exclusively be a matter of conflict between attitudes. If (1) you believe that you ought to  $p$  and (2) you believe that  $p$  only if you intend  $p$ , yet (3) you do not intend  $p$ , you are irrational. But there is no conflict among your attitudes; 1 and 2 do not conflict, and 3 cannot “conflict” with any attitudes because it is the absence of an attitude.

42 I consider a set with a present and an absent attitude to be a set with two attitudes\*.

43 It must merely be *possible* to generate the contradiction from  $M \cup S_M$ . Here is why. Suppose (1) you believe that  $2 = 1$ . The success proposition of this belief generates a contradiction. But you are not necessarily irrational. However, suppose (2) you believe that  $2 = 1$  and you believe that not- $(2 = 1)$ . In this case, the two beliefs’ success propositions generate the same contradiction as in 1. However, this time you are irrational. That is why there must merely be the *possibility* of generating a contradiction from  $M \cup S_M$ .

The consequence of this condition is clear. The fact that the attitudes in  $M$  cannot succeed jointly must be a consequence of how the attitudes\* in  $M$  relate to each other. Individual attitudes thus cannot be irrational. As wrongheaded as your belief that  $2 = 1$  or your intention to be married to a bachelor may be, it is not necessarily irrational.

I agree that no individual first-order attitude should make a person necessarily structurally irrational. Nevertheless, I disagree that we should make EISA subject to Multiple. There are two chief reasons for this. First, the condition goes too far. There are cases where it exempts you from being irrational but where no such exemption should be granted. Second, it does not go far enough. It fails to exempt you from being irrational where such an exemption should be granted.

Here is my first objection. It shows that Multiple goes too far. Suppose you believe that ( $p$  and not- $p$ ). (Note that this is a first-order belief with a non-atomic content.) Many philosophers view such an individual belief in a contradiction as structurally irrational. If this is correct, then Multiple turns out to be a nonstarter. The success proposition of such a belief does entail<sub>me</sub> a contradiction, yet this contradiction is not generated from at least two attitudes\*. Multiple thus precludes the irrationality of this belief, and there is no justification for that. This shows that Multiple does not state a prerequisite for structural irrationality.

Here is another aspect we should consider. I argue that making EISA subject to Multiple would be *ad hoc*. It does not help us to get to the core of structural irrationality. One and the same individual or object can figure in an attitude under various modes of representation. For example, one can represent the same person as “Batman” or “Bruce Wayne.” Likewise, one can represent one and the same planet as “Hesperus” or “Phosphorus.” Consider two examples where this becomes significant for structural irrationality. Suppose you believe that Batman is a hero and you believe that Bruce Wayne is not a hero. Or suppose you intend to observe Hesperus and you intend not to observe Phosphorus:

{<Belief: Batman is a hero>; <B: not-Bruce Wayne is a hero>;}

{<Intention: You observe Hesperus>; <I: not-You observe Phosphorus>;}.

First note that, according to EISA, these two sets are necessarily irrational. It is thus metaphysically impossible for Batman to possess a property that Bruce Wayne lacks. Likewise, it is metaphysically impossible to observe Hesperus without observing Phosphorus, and *vice versa*. Thus, the success propositions of both sets entail<sub>me</sub> a contradiction.

This reveals another defect of EISA. Subscribing to either of these two sets of attitudes does not suffice to make you irrational. You may blamelessly lack awareness that (Batman and Bruce Wayne) and (Hesperus and Phosphorus) are identical. The lack of awareness may not be a rational defect on your part. Again, EISA overshoots in ascribing irrationality.<sup>44</sup>

However, Multiple fails to provide an adequate solution to this problem. Multiple implies that irrationality can only arise for sets of attitudes\* if one can generate the impossibility of success from the relationship that holds among those attitudes\*. This is precisely the case for the two examples just discussed. Both represent *antagonistic* pairs of attitudes. Your belief that Batman is a hero succeeds if and only if your belief that Bruce Wayne is not a hero does not succeed. Your intention to observe Hesperus succeeds if and only if your intention not to observe Phosphorous does not succeed.

I conclude that Multiple turns out to be inadequate when it comes to adapting EISA appropriately. While it manages to exclude individual attitudes from being structurally irrational, it also excludes beliefs that are in flat contradiction from being structurally irrational. It also offers no solution for excluding attitudes that entail<sub>me</sub> a contradiction under various modes of representation from being structurally irrational. Multiple is not part of the essence of structural irrationality. We should not make EISA subject to Multiple.

### 5.2. Reflective Accessibility

I propose an alternative, unified solution to the problems discussed in the previous section. I argue that irrationality presupposes *reflective accessibility*, as I shall put it. I will stipulate that reflective accessibility puts three fundamental constraints on any correct account of irrationality. First, it requires *objective transparency*. That is, a set of attitudes\* is irrational only if it is objectively transparent that its members cannot succeed jointly when adopted. Second, it requires *subjective transparency*. That is, a set of attitudes\* is irrational only if the person who holds the attitudes is able to infer that they cannot succeed jointly when adopted. Third, it requires *implicit approval*. That is, a set of attitudes\* is irrational only if you are *not* justified in believing that the attitudes can succeed jointly when adopted. A correct modification of EISA must be sensitive to these three conditions, or so I shall argue in the following.

44 Here is a more general expression of this problem. Suppose  $p$  and  $q$  are incompatible in the following sense:  $p$  entails not- $q$ , and  $q$  entails not- $p$ . However, suppose you are not in a position to know that. That is, even if you were to consider  $p$  and  $q$  while employing your full conceptual and logical abilities, you could not come to discover that  $p$  and  $q$  cannot be true together. EISA would still imply that you are irrational. Your attitudes cannot succeed together, yet this result is clearly untenable.

## 5.2.1. Objective Transparency

In this section, I will only focus on objective transparency. In the previous section, I mentioned two types of problems for EISA. First, some individual attitudes turn out to be necessarily irrational. Second, there are sets of attitudes that, due to their contents' mode or sense of representation, are not necessarily irrational, even though they cannot succeed jointly. Both types of problems can be resolved by introducing an objective transparency condition.

The basic idea behind objective transparency is this. Consider a set of attitudes\* that cannot possibly succeed when adopted. Then, for these attitudes\* to be irrational, this impossibility must be objectively transparent to a certain degree.

I assume that the degree to which the impossibility of success is objectively transparent depends on the (complexity of the) attitudes\* and the inferential abilities one would have to deploy to ascertain that a set of attitudes\* cannot succeed jointly. This relation is, of course, inverse: the more (complex) the attitudes and abilities needed to ascertain that a set of attitudes\* cannot succeed jointly, the less objectively transparent the circumstances (and *vice versa*).

Thus defined, EISA presupposes an extremely low degree of objective transparency. In order to establish whether a set of attitudes is irrational, one needs to be able to ascertain virtually *all* metaphysically necessary falsities.<sup>45</sup> That is quite a tall order; no actual person satisfies this. Some metaphysical falsities may be too complicated for anyone to understand. Others may have yet to be discovered.<sup>46</sup> This exposes an elemental flaw in EISA. We should not accept an account of irrationality that allows for irrational combinations of attitudes that no actual person can identify as such. EISA presupposes insufficient objective transparency.<sup>47</sup>

45 For example, I explained that EISA implies that intending to observe Hesperus while intending not to observe Phosphorous is irrational. In order to discern that, one needs to have a substantial and rather complex piece of information. One needs to be aware that Hesperus = Phosphorous. That is quite a substantial requirement.

46 Here is a case in point: the axioms of arithmetic may either imply or contradict Goldbach's conjecture. So far, no one has settled the matter. Suppose, for the sake of argument, that the axioms of arithmetic contradict Goldbach's conjecture. Then, according to EISA, it would be irrational to believe both. But no one could actually show that this is the case (cf., Broome, *Rationality out of Reasoning*, 154).

47 By contrast, consider a condition of objective transparency that would be too strong. Suppose that for a set of attitudes\* to be irrational, the impossibility of success must be transparent to the degree of an explicit contradiction. By this I mean that for a set of attitudes  $M$  to be irrational, conjoining  $M$  with its success propositions  $S_M$  must lead to a set that contains two explicitly contradictory propositions. In short,  $M \cup S_M$  must contain " $p$ " and " $\text{not-}p$ ." This condition is untenably strong. Looking at the paradigmatic examples of

I stipulate that the correct degree of objective transparency is best designated as “logical transparency.” When considering a set of attitudes and their success propositions, one must be able to infer that not all attitudes in that set can succeed simultaneously *via a logically valid inference*. Here is a slightly more formal expression of this idea. Let  $M$  again be the propositional representation of your attitudes\*, and let  $S_M$  be their success propositions. Then, in order for an account to pick out the irrationality of the attitudes  $M$  represents correctly, the account must be subject to the following condition:

*Objective Transparency:*  $M \cup S_M$  entails a contradiction *qua* logically valid inference.

Again, by “contradiction” I mean a formal or explicit contradiction, i.e., a proposition and its negation ( $p$  and not- $p$ , for example). Moreover, it is critical to note here that I use “logically valid inference” in a rigid sense.

You may deem an inference logically valid if it necessarily preserves truth. In this sense, an inference from “Peter is unmarried” to “Peter is a bachelor,” or from “Mary is an ophthalmologist” to “Mary is a doctor,” would also be logically valid. Yet this is *not* my understanding of logical validity. I understand logical validity as strict formality or logical consequence. A logically valid inference is an inference the validity of which is entirely general, or topic neutral.<sup>48</sup> It preserves truth *qua* its logical form, not *qua* the substance or meaning of the proposition that constitutes it.<sup>49</sup> So, unlike the two abovementioned inferences regarding Peter and Mary, a logically valid inference remains valid even when abstracting from the particular meaning or semantic content of the objects that constitute it.<sup>50</sup>

I will rely on a rough-and-ready assessment to see whether an inference is logically valid. A logically valid inference remains strictly truth preserving even when one *anonymizes* the semantic content of the inference. For example, the inference “No fish is a mammal; some animals are mammals; so, some animals are not fish” satisfies this criterion. The inference remains truth preserving even when anonymized: “No  $F$  is  $M$ ; some  $A$  is  $M$ ; so, some  $A$  is not  $F$ .” The inference from “Peter is unmarried” to “Peter is a bachelor” does not remain

---

irrationality  $M_1$ – $M_6$ , only  $M_1$  and  $M_2$  would qualify as irrational. These are the only two patterns of attitudes where  $M \cup S_M$  contains two explicitly contradictory propositions.  $M \cup S_M$  of  $M_6$ , for example, reads as follows:  $\{p; p; \langle \text{not-intention: } q \rangle; q; \langle \text{intention: } q \rangle\}$ ; there is no explicit contradiction.

48 MacFarlane, “What Does It Mean to Say That Logic Is Formal?”

49 Cf. Beall and Restall, “Logical Consequence.”

50 Cf. MacFarlane, “What Does It Mean to Say That Logic Is Formal?”

truth preserving when anonymized:  $P$  is  $U$ ; so,  $P$  is  $B$ . This inference preserves truth not through its logical form but through a semantic understanding of its contents.<sup>51</sup>

Let us now turn to how objective transparency contributes to establishing a unified account of irrationality. It fixes some of the defects of EISA and avoids the complications of Multiple.

In the previous section, I identified two distinct flaws of EISA. EISA overreaches in implying the necessary irrationality of attitudes that are not necessarily irrational. Recall, for example, the two individual attitudes mentioned at the beginning of the previous section:

<Belief:  $2 = 1$ >;

<Intention: You are married to a bachelor>.

I have already shown how, in principle, Multiple can deal with this flaw. However, so can Objective Transparency. Consider the success propositions of your belief and intention:

< $2 = 1$ >;

<You are married to a bachelor>.

Though both propositions entail<sub>me</sub> a contradiction, they alone do not constitute a logically valid inference that entails a contradiction. To show this, conceal everything that belongs to the semantic content of these propositions and leave only their formal structure. Assuming that the numerical values 2 and 1 are part of the semantic content of  $2 = 1$ , an apt anonymization reads as the formal identity statement " $x = y$ ." Likewise, you may formalize "You are married to a bachelor" as "There exists one  $x$ :  $Mx$  and  $Bx$ ."

Trivially, neither the formal identity relation nor every existentially quantified conjunction implies a contradiction. The entailed contradiction is not a consequence of the formal structure of the statements. Neither statement implies a contradiction via a logical inference. Objective Transparency therefore corrects EISA in turning the attitudes in question into necessarily irrational ones.<sup>52</sup>

51 Anonymizing a proposition involves a number of complicated issues. Of course, finding a clear-cut demarcation between semantic content and logical form is not always straightforward. Moreover, anonymizing must also be sensitive to the logical form that comes with a subject's representation of a proposition.

52 This is not to say, of course, that by adopting either of these two attitudes you will not likely be irrational. I assume that most people believe that  $2 \neq 1$  and that you cannot be married

Let us now turn to the second critical problem of EISA. Recall the following set of attitudes:

{<Belief: Batman is a hero>; <Belief: not-(Bruce Wayne is a hero)>}

{<Intention: You observe Hesperus>; <Intention: not-(You observe Phosphorus)>}

Since it is metaphysically impossible for your beliefs and intentions to succeed jointly, EISA picks them out as irrational.

I have already explained that Multiple is toothless when it comes to fixing the problem. Nevertheless, Objective Transparency manages to deal with it. Consider the success propositions of the attitudes just mentioned:

<Batman is a hero>; <not-(Bruce Wayne is a hero)>.

<You observe Hesperus>; <not-(You observe Phosphorus)>.

I treat “Batman is a hero” and “Bruce Wayne is a hero,” as well as “You observe Phosphorous” and “You observe Hesperus,” as two distinct propositions. Following Kripke, I take this to be a precondition for the possibility of your being able, for example, to intend (or believe) the one proposition without intending (or believing) the other.<sup>53</sup> Indeed, I take this to be a distinct possibility. It follows from the following relationship between accepting and believing:

If an agent *A* sincerely, reflectively, and competently accepts a sentence *s* . . . , then *A* believes, at the time of *c*, what *s* expresses in *c*.<sup>54</sup>

I assume that one can “sincerely, reflectively, and competently” accept “You observe Phosphorus” while not accepting “You observe Hesperus.” Thus, they are two distinct propositions.

This allows us to anonymize the success propositions of the patterns of attitudes above as follows:

<*p*>; <not-*q*>;

<*r*>; <not-*s*>.

---

to a bachelor. By adopting one of the attitudes in question, you will be subscribing to an irrational set of attitudes.

53 “It also seems clear that there must be two distinct propositions or contents expressed by ‘Cicero denounced Catiline’ and ‘Tully denounced Catiline.’ How else can Tom believe one and deny the other? And the difference in propositions thus expressed can only come from a difference in sense between ‘Tully’ and ‘Cicero’” (Kripke, “A Puzzle about Belief,” 243). Cf. also McKay and Nelson, “Propositional Attitude Reports.”

54 McKay and Nelson, “Propositional Attitude Reports.”



This makes it clear why, subject to Objective Transparency, EISA no longer implies the irrationality of the above attitudes. None of these pairs of success propositions licenses a logically valid inference to a contradiction. Given the absence of the insight that (Batman = Bruce Wayne) and (Hesperus = Phosphorus), no irrationality is on display here. So even if these attitudes cannot jointly metaphysically succeed, this is not sufficiently transparent to give rise to structural irrationality.

Unlike Multiple, Objective Transparency does not go too far here. It fixes the problem of assigning two attitudes, the incompatibility of which is disguised by the different senses of the attitudes' contents. This already suggests that Objective Transparency is preferable to Multiple as an appropriate weakening of EISA. There is further evidence that this is true, however. Recall that Multiple necessarily exempts a belief (or intention) in an explicit contradiction from being irrational. I have already indicated that such an exception can be granted in extraordinary circumstances (which I will define more closely in the next section). However, such an exception is far from *necessary*.

Objective Transparency does not imply a necessary exemption. Consider the anonymized success proposition of a belief in or intention to realize an explicit contradiction:

< $p$  and not- $p$ >.

Trivially, this entails a contradiction *qua* a logically valid inference. Thus, the requirement of *Objective Transparency* is consistent with the idea that single intentions and beliefs in explicit contradictions are attitudinally irrational. Objective Transparency, again, turns out to be superior to Multiple.

So far, I have shown three things. First, EISA goes too far in assigning irrationality to individual attitudes and combinations of attitudes with contents that differ in their mode of representation. Second, I have shown that the traditional solution to this problem, Multiple, is also inadequate. Multiple manages to avoid portraying individual attitudes as irrational, but it fails to deal appropriately with beliefs or intentions in flat contradictions and with attitudes that take different (or Fregean) attitudes toward one and the same thing or subject. Third, Objective Transparency turns out to be superior to Multiple. When considering the attitudes and their success propositions, the impossibility of success must be transparent to the degree of a logically valid inference. I suggested that this weakens EISA in the right way.

In order to satisfy Objective Transparency, EISA must be slightly adapted. I will call the adapted account the *Formal Impossible Success Account* (FISA). It reads as follows:

*FISA Formal*:  $M$  represents an irrational set of attitudes\* if and only if  $M \cup S_M$  entails<sub>fo</sub> a contradiction,

where “entails<sub>fo</sub>” means “entails formally.” I use “entails formally” as shorthand for “entails *qua* logically valid inference.”

By replacing “entails<sub>me</sub>” with “entails<sub>fo</sub>” in EISA, we make the resulting account sensitive to Objective Transparency. The fact that an irrational set of attitudes makes joint success impossible must be transparent to the degree of a valid inference.

I have already shown that this adaptation manages to eliminate some of EISA’s and Multiple’s critical flaws. But how does it fare with regard to the irrationality of  $M_1$ – $M_6$ ? Can FISA still explain their irrationality? To begin, FISA straightforwardly preserves the irrationality of  $M_1$ – $M_2$  and  $M_4$ – $M_6$ . For  $M_1$  and  $M_2$ , this is virtually trivial; I will not demonstrate this here. As a proxy for  $M_4$ ,  $M_5$ , and  $M_6$ , consider the union of  $M$  and  $S_M$  for  $M_5$ :

$$\{p; p \Rightarrow \langle \text{intention: } q \rangle; \langle \text{not-intention: } q \rangle\}.$$

This set also entails<sub>fo</sub> a contradiction:  $p$  and  $p \Rightarrow \langle \text{intention: } q \rangle$  entails<sub>fo</sub>  $\langle \text{intention: } q \rangle$ , which, when conjoined with  $\langle \text{not-intention: } q \rangle$ , entails a contradiction.

There is one outlier, however. Suppose you hold a set of cyclical preferences, as represented by  $M_3$ . Suppose these preferences are based on a set of conditional intentions. As discussed in section 2, I assume that a conditional intention to ( $a$  and not- $b$ ) if  $a$  or  $b$  is successful only if not- $b$  is true. Accordingly, this generates the following set of success propositions:

$$\{\text{not-}b; \text{not-}c; \text{not-}a\}.$$

According to EISA (and its predecessor SISA), there are situations in which these preferences turn out to be irrational. This is precisely the case if, and only if,  $a$ ,  $b$ , and  $c$  are necessarily contrary propositions, i.e., it is impossible for all of them to be false. Then the three conditionals entail<sub>me</sub> a contradiction. According to FISA, however, this is no longer the case. The three propositions do not entail<sub>fo</sub> a contradiction. A set of conditional intention–based and cyclical preferences does not qualify for irrationality. Should we accept this?

In general, I believe we should. Suppose you have a cyclical set of preferences that is constituted by conditional intentions. Suppose you are innocuously unaware that there is no metaphysically possible world at which  $a$ ,  $b$ , and  $c$  are collectively false. Perhaps you even justifiably believe that you can avoid failing to satisfy your conditional preferences simultaneously. Then you are not necessarily irrational. Moreover, FISA still permits a clear-cut scenario where a

cyclical set of preferences makes you attitudinally irrational. As I explained in section 2, cyclical preferences of that kind make you irrational whenever you hold them in conjunction with a conditional intention-based preference to realize one of the options (i.e.,  $a$ ,  $b$ , or  $c$ ) rather than none of the options (i.e., not- $[a, b, \text{ and } c]$ ). The same is true if you also believe that it is impossible for none of the options to be implemented. As I explained in section 2, both your belief and your preference take “ $a$  or  $b$  or  $c$ ” as a success proposition. So, by adding either this preference or this belief, the set of success propositions for your attitudes reads as follows:

$$\{\text{not-}b; \text{not-}c; \text{not-}a; a \text{ or } b \text{ or } c\}.$$

This set entails<sub>f<sub>o</sub></sub> a contradiction. First, the first three propositions are jointly true only if  $a$ ,  $b$ , and  $c$  are jointly false. However, the truth of the remaining success proposition depends on  $a$ ,  $b$ , and  $c$  *not* being jointly false. Consequently, this set entails<sub>f<sub>o</sub></sub> a contradiction. FISA can thus account for the irrationality of cyclical conditional intention preferences. But this is only the case if either you have a preference to realize at least one of these options or you have a belief that makes it objectively transparent to you (as it were) that your preferences cannot succeed jointly.

Let us now turn to cyclical preferences that are based on comparative desires. As discussed in section 2, I assume that a comparative desire-based preference for  $a$  over  $b$  succeeds only if  $a$  is (in some sense) *Fer* than  $b$ . This generates the following set of success propositions  $S_{M_3}$ :

$$\{a \text{ is } Fer \text{ than } b; b \text{ is } Fer \text{ than } c; c \text{ is } Fer \text{ than } a\}.$$

I have already explained that these propositions entail<sub>me</sub> a contradiction. The “\_\_\_ is *Fer* than \_\_\_” relation is necessarily transitive. Cyclical preferences that are based on comparative desires are thus irrational under EISA. But are they irrational under FISA as well? This poses an interesting question, for it depends on whether you consider the “\_\_\_ is *Fer* than \_\_\_” relation to be part of the semantic content or part of the formal structure of “ $a$  is better than  $b$ .” Suppose it is entirely part of the semantic content. Then proper anonymizing would need to mask the semantic elements of “better than.” We would need to represent “ $a$  is *Fer* than  $b$ ” as

$$aRb,$$

only conveying that  $a$  relates to  $b$ . The anonymized set of success propositions would thus look as follows:

$$\{aRb; bRc; cRa\}.$$

This does not entail<sub>f<sub>0</sub></sub> a contradiction. Like the example above, this set entails<sub>f<sub>0</sub></sub> a contradiction only if we add an attitude with the success proposition “( $aRb$  and  $bRc$ )  $\rightarrow$  not- $cRa$ ” to the above set. An example is the belief that if  $a$  is *Fer* than  $b$  and  $b$  is *Fer* than  $c$ , then  $c$  is not *Fer* than  $a$ . Together with this belief, a set of desire-based preferences would turn out to be irrational.

Alternatively, we could suppose that at least the comparative part of the betterness relation belongs to the logical structure of the proposition “ $a$  is *Fer* than  $b$ .” Then “ $a$  is *Fer* than  $b$ ” already represents the anonymized proposition. The success propositions of desire-based cyclical preferences would read as follows:

{ $a$  is *Fer* than  $b$ ;  $b$  is *Fer* than  $c$ ;  $c$  is *Fer* than  $a$ }.

Arguably, this entails<sub>f<sub>0</sub></sub> a contradiction, at least if we treat the inferences from “ $a$  is *Fer* than  $b$ ” and “ $b$  is *Fer* than  $c$ ” to “ $a$  is *Fer* than  $c$ ,” as well as from “ $a$  is *Fer* than  $c$ ” to “not-( $c$  is *Fer* than  $a$ )” as formal—an assumption I do not find entirely implausible. As a result, a set of desire-based cyclical preferences would turn out to be irrational on its own.

### 5.2.2. Subjective Transparency

So far, I have defended the following picture of structural irrationality. For a set of attitudes\* to be irrational, holding those attitudes\* must preclude the joint success of the present attitudes in that set. But that is not enough. The mentioned impossibility must also be *reflectively accessible*. It must be entailed via a formally valid inference by the propositions that represent the attitudes and their success conditions.

In this section, I will add another (related but more subjective) condition. I assume that irrationality presupposes the *ability* to detect an error in one’s attitudes. That is, you must be able to see that an irrational set of attitudes is defective. Suppose you hold two contradictory intentions. You must have sufficient logical and inferential abilities to see that you have arranged your attitudes inadequately.

It is crucial to note that I am only stipulating a counterfactual condition here. For a set of attitudes\* to be irrational, there is no need to actually identify a necessary defect. That would be a blatant way of “over-intellectualizing” irrationality. Instead, the idea is that if you were to entertain an irrational set of attitudes, you would come to see that something was amiss.

Here is the general idea I have in mind. Suppose you intend to  $p$  and believe ( $p$  only if you intend  $q$ ), yet you do not intend  $q$ . Suppose you call these attitudes\* to your consciousness. So you say to yourself “I will  $p$ , but only if I intend  $q$ . But I do not intend  $q$ .” If, upon employing your full logical and inferential capacities, you were to conclude that your attitudes are *not* incompatible, then

I suggest that you are indeed exempt from being irrational. In other words, you are irrational only if you have the capacity to infer that your attitudes are somehow defective or incompatible.<sup>55</sup>

The hardest thing in this context is to pin down exactly what kind of logical and inferential capacity irrationality presupposes. My suggestion is that the required capacity will coincide with the logical capacity to infer a contradiction from a set of attitudes and their success propositions. That is, suppose that a pattern of attitudes precludes those attitudes' joint success and that this is transparent to the degree of a valid inference. By adopting this pattern of attitudes, you are irrational only if you have the capacity to infer the contradiction that is entailed by your attitudes\* and their success conditions. Put succinctly: you must be able to deduce a contradiction from  $M \cup S_M$ . This is the sense, I assume, in which irrationality presupposes the *ability* to detect a flaw in your attitudes\*.

Before I add this condition to FISA, let me anticipate, albeit briefly, two possible critiques. On the one hand, you may think that this condition is too weak. As I envision it, the condition says that if you were to call a set of your attitudes\* and their success propositions to your cognitive attention while entertaining your maximal inferential capacity, you would infer a contradiction. You may think that this counterfactual should be restricted further. Perhaps you think it should only include those attitudes to which you have conscious access.

The effect of that restriction would be clear. For a set of attitudes to be irrational, one needs to be able to consciously entertain them.<sup>56</sup> I believe that would be overly restrictive, however. For one, it would preclude modelling "unconscious or implicit biases" (i.e., unconscious or automatic prejudicial attitudes that predicate your social behavior) as structural irrationality. We should not deprive a theory of irrationality of that possibility. On the other hand, you may think that the condition is too strong. Suppose you intend  $p$  and believe that  $p$  only if you intend  $q$ , yet you do not intend  $q$ . Also, if you were to consider your attitudes\* and their success propositions, you would infer a contradiction. Yet you explicitly deny that truth and implementation are success conditions for belief and intention, respectively. Should you not be exempt from being irrational?

My quick answer to this is no. The conditions for irrationality must be subjective, but not too subjective. In fact, Broome has already established this point

55 Of course, there is an important caveat. The error you identify needs to relate to the constitutive aims of your attitudes. It is not enough to conclude, for example, that combining two attitudes is phenomenologically or aesthetically displeasing or incompatible.

56 Cf. Wallace, "Normativity, Commitment, and Instrumental Reason," 22.

convincingly.<sup>57</sup> Following this point and not repeating the argument here, I do not think that a theory of (ir)rationality should be predicated, by and large, on your subjective sensitivities toward success conditions. Moreover, in some implicit sense, I doubt that you can detach yourself from the success conditions of your attitudes. Success conditions are grounded in essential dispositions that come with your attitudes. If you competently judge that you ought to  $p$ , then, for as long as you are aware that it is not the case that you ought to  $p$  or that  $p$  is impossible, you will be disposed\* to discard your judgment. This is the sense in which I suppose that the success of your attitudes plays a correcting or structuring role. You will normally discard an attitude when you become aware that its success conditions will not be satisfied. So at least in a dispositional or implicit sense, I take it that you will subscribe to the success conditions of your attitudes. It is this implicit adherence that makes it plausible to include success propositions as part of an inference that you need to be able to perform in order to count as irrational.

In order to make FISA subject to this condition, I shall qualify the account as follows. I will refer to the resulting account as FISA<sup>+1</sup>:

FISA<sup>+1</sup> *Formal*:  $M$  represents an irrational set of attitudes\* if and only if (1)  $M \cup S_M$  entails<sub>f<sub>0</sub></sub> a contradiction  $C$  and (2) you can infer  $C$  from  $M \cup S_M$ .

As mentioned above, it is important to emphasize the counterfactual character of “can infer.” Irrationality does not presuppose that you must actually infer a contradiction by considering  $M \cup S_M$ . That would be a blatant case of over-intellectualizing one’s account of irrationality. Rather, it presupposes that if you were made aware of the propositions in  $M \cup S_M$  and you were to utilize your full logical capacities (i.e., making all the correct inferences you can make), you would infer a contradiction.

### 5.2.3. *Implicit Approval*

Finally, let us turn to the third condition of reflective accessibility: *implicit approval*. This condition says that if your attitudes\* and success propositions entail<sub>f<sub>0</sub></sub> a contradiction, you are irrational only if you *implicitly accept* the validity of the formal entailment and the falsity of the entailed contradiction. The acceptance must only be implicit insofar as it only requires the *absence* of a justified belief. I will explain this below.

57 Broome, *Rationality through Reasoning*, sec. 6.2: “Your rationality cannot be judged entirely by your own standards” (93).

An example inspired by the so-called preface paradox illuminates the necessity of this condition. Suppose that, on the basis of good evidence, you believe that every single proposition in a book you have authored is true. Suppose, again on the basis of good evidence, that you also believe there is no book for which it holds that every statement is true. Clearly, the success propositions of your beliefs entail<sub>f<sub>0</sub></sub> a contradiction. According to FISA<sup>+1</sup>, this combination of beliefs turns out to be irrational (given that you can infer a contradiction from the two success propositions).

However, suppose you are also an accomplished dialetheist (I am thinking of someone like Graham Priest here). You have developed a precise view as to when a contradiction turns out to be true. You agree that (1) “Every single proposition in my book is true” and (2) “No book contains only true propositions” entails a contradiction, although you do not accept that the truth of 1 excludes the truth of 2 (and *vice versa*). From your considered standpoint, both of your beliefs can succeed at the same time. Are you irrational?

I would say no. You have constructed a refined argument from which it follows that both beliefs can succeed. You rely on an arsenal of grounds for why this is so. I assume you are *justified* in believing that your two beliefs can succeed simultaneously. So, if the world is as you justifiably take it to be, it is possible for your attitudes to succeed simultaneously. It is not credible to deem you irrational.<sup>58</sup>

Here is an actual case where this point becomes relevant. Suppose we treat the inference from “*a* is Fer than *b*” and “*b* is Fer than *c*” to “*a* is Fer than *c*” as a formal entailment (I discussed an alternative to this view in section 5.2.1.) Suppose, however, that on the basis of your evidence, you justifiably deny the necessary transitivity of comparative “\_\_\_ is Fer than \_\_\_” relations (think of someone like Larry Temkin here). That is, you deny the validity of the inference from “*a* is Fer than *b*” and “*b* is Fer than *c*” to “*a* is Fer than *c*.” If your considered view is correct, then a set of cyclical preferences can succeed together. An ascription of irrationality would not be warranted.

In order to make FISA<sup>+1</sup> subject to these considerations, I suggest extending the account by adding a third condition (3) to it.

*FISA<sup>+2</sup> Formal:* *M* is irrational if and only if (1)  $M \cup S_M$  entails<sub>f<sub>0</sub></sub> a contradiction *C*, (2) you can infer *C* from  $M \cup S_M$ , and (3) you have neither (a) a justified belief that the entailment from  $M \cup S_M$  to *C* is invalid nor (b) a justified belief that *C* is true.

58 Cf. Broome, *Rationality through Reasoning*, 91.

This added condition ensures that if you have reached a robust vantage point from which it is possible for a set of attitudes to succeed at the same time, it is no longer defensible to accuse you of irrationality, even if your view turns out to be incorrect. Recall that structural irrationality is tied up with subjective incoherence. Therefore, if a belief of yours that all attitudes\* in a set *M* can succeed simultaneously is well founded and epistemically justified, we have no grounds to accuse you of being irrational in adopting *M*.

It is important not to conceive of this as an easy excuse. I will not adopt a particular position on epistemic justification here. However, this kind of justification must satisfy a few important constraints. First, justification must be an entirely internal matter and must supervene on the mind. Otherwise, my account will violate the rule that structural rationality supervenes on the mind.<sup>59</sup> Second, justification must be in some sense rigorous and demanding. Being justified in believing that a particular set of attitudes can be jointly successful is not something you can bootstrap into existence. So, an unfounded belief that you are justified will not suffice. The bar for justification must be set significantly higher. In fact, I imagine that only a handful of specialized philosophers and Buddhists (perhaps only Graham Priest!) satisfy this condition.

## 6. CONCLUSION

This paper has sought to establish the essence of structural irrationality. I aimed to determine what creates and unifies the domain of structurally irrational attitudes. In a nutshell, I argued that structural irrationality consists in the transparent suspension of the possibility of attitudinal success. The core principle of this account can be formulated as follows: a set of attitudes\* is irrational if and only if holding those attitudes\* precludes the joint success of the present attitudes in that set. I formalized this by assuming that every attitude comes with a set of success propositions. If that set, conjoined with the propositions that represent the present and absent attitudes, entails a contradiction, then we have identified an irrational set of attitudes.

Although the core principle alone comes with considerable explanatory power—in sections 3 and 4, I showed that it can unify six fundamental types of structural irrationality—I also argued that not every suspension of the possibility of attitudinal success instantiates a set of irrational attitudes. If, for example, you believe or intend a necessary falsity, then your attitudes cannot succeed, but you are not necessarily irrational. I argued that the suspension must be reflectively accessible. I have defined three conditions to guarantee this.

59 Broome, *Rationality through Reasoning*, 89, 151–52; Wedgwood, “Internalism Explained,” 349.



First, the suspension of the possibility of success must be objectively transparent. That is, by looking at the attitudes\* and their success propositions, it must be transparent to the degree of a formally valid inference that the attitudes in question cannot succeed jointly. Second, and related, you must have the logical ability to perform that inference. That is, by looking at the attitudes\* and their success propositions and by mustering all of your logical and conceptual abilities, you must be able to discern that it is impossible for your attitudes to succeed. Third, you must not be justified in believing that the attitudes can succeed jointly or that the inference in fact lacks the property of being truth preserving. If these three conditions are met, then a set of attitudes turns out to be irrational.

As a consequence of this proposal, the following picture of irrationality emerges. Irrationality consists in a constitutive defect with regard to your attitudes. This defect stems entirely from how you have structured your attitudes\*. In addition, you are able to discern the defect. Of course, whether this account can capture the full range of phenomena that are vulnerable to structural irrationality, which arguably includes certain combinations of graded belief/credence as well as combinations involving attitudes such as hopes, fears, and the like, is something future research will need to show.

As a final thought, suppose that the constitutive success of agency is invariably linked to the success of an agent's attitudes. Then the picture of rationality I have drawn in this paper is one in which structural rationality can be seen as the most foundational prerequisite for being a successful agent.<sup>60</sup>

University of Bayreuth  
julian.fink@uni-bayreuth.de

#### REFERENCES

- Anscombe, G. E. M. *Intention*. Cambridge, MA: Harvard University Press, 1957.
- Beall, Jc, and Greg Restall. "Logical Consequence." In *Stanford Encyclopedia of Philosophy* (Winter 2016). <https://plato.stanford.edu/archives/win2016/entries/logical-consequence/>.
- Broome, John. "Enkrasia." *Organon F* 20, no. 4 (2013): 425–36.
- . *Ethics out of Economics*. Cambridge: Cambridge University Press, 1999.
- . *Rationality through Reasoning*. Chichester: Wiley Blackwell, 2013.

60 I am grateful to John Broome, Benjamin Kiesewetter, Franziska Poprawe, Andrew Reisner, Olivier Roy, Vuko Andrić, and numerous anonymous referees for comments on earlier drafts of this paper.

- . *Weighing Lives*. Oxford: Oxford University Press, 2004.
- Brunero, John. *Instrumental Rationality: The Normativity of Means-Ends Coherence*. Oxford: Oxford University Press, 2020.
- Fink, Julian. "A Constitutive Account of 'Rationality Requires.'" *Erkenntnis* 79, no. 4 (August 2014): 909–41.
- . "Structural Irrationality Does Not Consist in Having Attitudes You Ought Not to Have: A New Dilemma for Reasons-Violating Structural Irrationality." *Australasian Journal of Philosophy* (forthcoming).
- . "What (In)Coherence Is Not." *Grazer Philosophische Studien* 99, no. 2 (October 2022): 125–34.
- Hume, David. *A Treatise of Human Nature*. Oxford: Oxford University Press, 2000.
- Katsafanas, Paul. *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press, 2013.
- Kiesewetter, Benjamin. *The Normativity of Rationality*. Oxford: Oxford University Press, 2017.
- Kolodny, Niko. "Why Be Rational?" *Mind* 114, no. 455 (July 2005): 509–63.
- Kripke, Saul A. "A Puzzle about Belief." In *Meaning and Use*, edited by Avishai Margalit, 239–83. Dordrecht: D. Reidel, 1979.
- Lord, Errol. *The Importance of Being Rational*. Oxford: Oxford University Press, 2018.
- . "What You're Rationally Required to Do and What You Ought to Do." *Mind* 126, no. 504 (October 2017): 1109–54.
- MacFarlane, John. "What Does It Mean to Say That Logic Is Formal?" PhD diss., University of Pittsburgh, 2000. <http://johnmacfarlane.net/dissertation.pdf>.
- McKay, Thomas, and Michael Nelson. "Propositional Attitude Reports." In *Stanford Encyclopedia of Philosophy* (Spring 2014). <https://plato.stanford.edu/archives/spr2014/entries/prop-attitude-reports/>.
- Rabinowicz, Wlodek. "Modeling Parity and Incomparability." In *Patterns of Value: Essays on Formal Axiology and Value Analysis*, vol. 2, edited by Wlodek Rabinowicz and Toni Rønnow-Rasmussen, 201–28. Lund: Lund University, 2004.
- . "Value Relations." *Theoria* 74, no. 1 (March 28, 2008): 18–49.
- Reisner, Andrew. "Is the Enkratic Principle a Requirement of Rationality?" *Organon F* 20, no. 4 (2013): 436–62.
- Scanlon, Thomas. "Structural Irrationality." In *Common Minds: Themes From the Philosophy of Philip Pettit*, edited by Geoffrey Brennan, Robert Goodin, Frank Jackson, and Michael Smith, 84–103. Oxford: Clarendon Press, 2007.
- Shah, Nishi. "How Action Governs Intention." *Philosophers' Imprint* 8, no. 5

- (July 2008): 1–19.
- Temkin, Larry S. “Intransitivity and the Mere Addition Paradox.” *Philosophy and Public Affairs* 16, no. 2 (Spring 1987): 138–87.
- Wallace, R. Jay. “Normativity, Commitment, and Instrumental Reason.” *Philosophers’ Imprint* 1, no. 3 (December 2001): 1–26.
- Wedgwood, Ralph. “Internalism Explained.” *Philosophy and Phenomenological Research* 65, no. 2 (September 2002): 349–69.
- . *The Nature of Normativity*. Oxford: Clarendon Press, 2007.
- Worsnip, Alex. *Fitting Things Together: Coherence and the Demands of Structural Rationality*. Oxford: Oxford University Press, 2021.
- . Review of *The Normativity of Rationality*, by Benjamin Kiesewetter. *Notre Dame Philosophical Reviews*, July 2, 2018. <https://ndpr.nd.edu/reviews/the-normativity-of-rationality>.
- . “What Is (In)Coherence?” In *Oxford Studies in Metaethics*, vol. 13, edited by Russ Shafer-Landau, 184–206. Oxford: Clarendon Press, 2018.