Non-Naturalist Moral Realism, Autonomy and Entanglement

Graham Oddie

Topoi
An International Review of Philosophy

ISSN 0167-7411

Topoi
DOI 10.1007/s11245-017-9454-z

Springer

Topoi
DOI 10.1007/s11245-017-9454-z

Non-Naturalist Moral Realism, Autonomy and Entanglement

Graham Oddie[1]

Abstract It was something of a dogma for much of the twentieth century that one cannot validly derive an ought from an is. More generally, it was held that non-normative propositions do not entail normative propositions. Call this thesis about the relation between the natural and the normative Natural-Normative Autonomy (or Autonomy for short). The denial of Autonomy involves the entanglement of the natural with the normative. Naturalism entails entanglement—in fact it entails the most extreme form of entanglement—but entanglement does not entail naturalism. In a ground-breaking paper "The autonomy of ethics" Arthur Prior constructed some intriguing counterexamples to Autonomy. While his counterexamples have convinced few, there is little agreement on what is wrong with them. I present a new analysis of Autonomy, one which is grounded in a general and independently plausible account of subject matters. While Prior's arguments do establish shallow natural-normative entanglement, this is a consequence of simple logical relationships that hold between just about any two subject matters. It has nothing special to do with the logical structure of normativity or its relation to the natural. Prior's arguments (along with several others) leave the fundamental idea behind natural-normative Autonomy intact. I offer a new argument for deep entanglement. I show that in any framework adequate for dealing with the natural and the normative spheres, a purely natural proposition entails a purely normative proposition, and vice-versa. But this is no threat to non-naturalist moral realism. In fact it helps ameliorate the excesses of an extreme non-naturalism, delivering a more palatable and plausible position.

Keyword Non-naturalist normative realism · Naturalism · Realism · Autonomy of ethics · Entanglement · Arthur Prior

1 Autonomy and Naturalism

It was something of a dogma for much of the twentieth century that one cannot derive an *ought* from an *is*. More generally, it was held that one cannot derive normative propositions from natural propositions. Call this thesis about the relation between the natural and the normative Natural-Normative Autonomy (Autonomy for short). The denial of Autonomy involves the *entanglement* of the natural with the normative. Autonomy has its historical roots in a famous observation in Hume, and received a strong boost at the beginning of the last century from Moore's influential Open Question Argument.

There are different varieties of moral realism but all of them hold that there are moral facts. This can be broken down into a combination of theses. The first is that there are genuine, non-trivial, moral propositions. (Non-cognitivists deny this.) The second is that some of these non-trivial moral propositions are true. (Nihilists and error theorists concede the first but deny the second.) The third is that non-trivial moral truths are made true by (perhaps *inter alia*) genuine moral facts.

Naturalists hold that, at bottom, there are only natural facts—that a complete specification of the purely natural facts is a complete specification of the world. This does not amount to a denial of genuine moral facts, just as physicalism does not amount to a denial of mental facts. But the

✉ Graham Oddie
oddie@colorado.edu

[1] Philosophy Department, University of Colorado at Boulder, UCB 232, Boulder, CO 80309-0232, USA

Published online: 25 March 2017

naturalist does contend that if there are any moral facts they must all, at bottom, be nothing but natural facts, since all the facts are purely natural facts. So naturalism is committed to the thesis that all the moral facts boil down to purely natural facts.

Suppose that naturalism is true, that every moral fact (if there are any) is a purely natural fact. The moral realist who embraces naturalism would then seem hard pressed to accept Autonomy. For let O be some true, non-trivial moral proposition. Since O is true and non-trivial, it is made true by the obtaining of some genuine fact F, and, since naturalism is true, F has to be a purely natural fact. Now consider the proposition N^F—*that the purely natural fact F obtains*. Since N^F affirms the obtaining of a natural fact, it is (presumably) a natural proposition. Since F is what makes O true, and N^F affirms the obtaining of F, it is not possible for N^F to be true and O false.[1] So a natural proposition—N^F—entails a non-trivial moral truth O. That is, Autonomy is false.[2]

This is, at best, a plausibility argument for the existence of counterexamples to Autonomy, but it does not furnish a concrete counterexample. It would be better if Autonomy skeptics were to exhibit at least one plausible, concrete counterexample to Autonomy. In his ground breaking paper, "The autonomy of ethics", Arthur Prior aims to do just this.[3] Many find Prior's counterexamples unconvincing for a variety of reasons. However, they do establish what I will call *shallow* natural-normative entanglement. What Prior's examples illustrate is the entanglement of normative and natural *content* in certain propositions, but they fall short of demonstrating entanglement of natural and normative facts. After outlining a new analysis of Autonomy, I show that shallow content-entanglement is a consequence of logical relationships that hold between a wide range of subject matters, even those that satisfy the guiding intuition behind Autonomy. I then offer a new argument against Autonomy, for what I call *deep* entanglement.

This conclusion is not in itself a threat to a non-naturalist moral realism. Naturalism does entail deep entanglement—indeed it entails the most extreme form of deep entanglement—but deep entanglement does not entail naturalism. The naturalist claims that every fact, including every normative fact, is a natural fact. The non-naturalist can happily deny this while countenancing deep, but nevertheless limited, natural-normative entanglement.

2 Two Versions of Autonomy: Entailment and Derivability

Autonomy has been characterized in a number of different ways. It is sometimes characterized as the thesis that no natural proposition *entails* a normative proposition. And it is sometimes characterized as the thesis that no normative proposition can be *derived* (or *inferred*) from a natural proposition. Call these the *entailment* and *derivability* versions of Autonomy respectively. Further, there are rival formulations of both theses in terms of entailment/derivability relations amongst *sentences* or *statements*, rather than the corresponding relations amongst propositions. What precisely is our target here?

I take entailment to be, in the first instance, a relation amongst propositions, and I assume a fairly standard account of this relation: propositions $P_1,\ldots, P_n$ *entail* proposition Q (abbreviated to: $P_1,\ldots, P_n \vDash Q$) if and only if every possible circumstance (or world-time) at which $P_1,\ldots, P_n$ are all true, Q is also true. Entailment amongst propositions can be carried over straightforwardly to entailment amongst sentences, provided the sentences at issue are interpreted—that is, they express propositions.[4] Because of this we can, without loss of generality, focus on propositions rather than sentences or statements. However, note that I do not assume that logically equivalent propositions are identical, or that propositions just *are* classes of world-times or functions from world-times to truth values. Propositions may be *structured* entities—hyperintensional rather than merely intensional entities—and so distinct propositions may well be logically equivalent.[5] Classes of world-times, or functions from world-times to truth values, are clearly not structured entities.

For our purposes here we need not go into the exact nature of structured propositions. All we will need is that each proposition *induces* a unique mapping from world-times to truth values. Where P and Q are propositions, the propositions expressed by the constructions $[P\vee Q]$ and $[\neg[P\wedge\neg Q]]$ can be distinct structured entities, but they

[1] This of course assumes that facts necessitate whatever propositions they make true. This is widely even if not universally endorsed.

[2] This is by no means intended as a watertight argument. Some naturalists have offered more detailed arguments for the systematic violation of Autonomy—e.g. (Jackson 2003, pp. 562–3). I think Jackson's argument is also flawed. The concrete counterexample to Autonomy that I construct here does not assume naturalism.

[3] Prior (1960).

[4] I am not particularly interested here in entailment relations amongst *uninterpreted* sentences. We can, of course, entertain the validity of certain *schemas*, involving variables ranging over interpreted sentences. A schema (such as $S\wedge T \vDash T\wedge S$) is valid if every substitution instance of the schema is valid.

[5] It may be that a perspicuous language is one in which the structure of a sentence perfectly mirrors the structure of the proposition it expresses, thus reducing the import of the sentence/proposition distinction. See (Tichý 1988) for an extended treatment of this.

induce the very same mapping from world-times to truth values. I will call the mapping, from world-times to truth values that the proposition P induces, the *content* of P (abbreviated: $P^{\mathbf{CON}}$).[6] Proposition P entails proposition Q just in case every world-time $P^{\mathbf{CON}}$ maps to true $Q^{\mathbf{CON}}$ does too.[7] Finally, a function that maps every world-time to the same value we call *trivial*. So a proposition has trivial content if it is necessarily true, or necessarily false or is necessarily undefined.

Unlike entailment, the notion of *derivability* is relative to some system D of rules of derivation. An entailment $P_1,\ldots,P_n \vDash Q$ is derivable in D (abbreviated to: $P_1,\ldots,P_n \vdash_D Q$) if one can demonstrate the validity of $P_1,\ldots,P_n \vDash Q$ by means of the D-rules. The entities dealt with in a derivational system have to be structured. This is so whether one thinks of the steps of a derivation as propositions, each of which is inferred by means of the rules of derivation from its predecessor, or whether one thinks of each step of a derivation as a *sequence* of propositions, each of which states an entailment (as in a Gentzen-style system). The rule of *disjunction introduction* for example, traffics in entities in which *disjunction* features as a *constituent*. But no function from world-times to truth values contains a truth function, or any other function, as a constituent.[8] Rules of derivation operate on essentially structured entities, rather than on their contents. So while entailment is a coarse-grained notion (*viz*. propositions with the same contents bear the same entailment relations) derivability may well be a fine-grained notion. It may be that $A \vDash C$, A and B have the same content, but $B \nvdash_D C$.

What we clearly want from a system of derivation is, first, that it only yields derivations of valid entailments (it should be sound); and secondly, that it yields derivations of all valid entailments (it should be complete). If D is both sound and complete then $P_1,\ldots,P_n \vDash Q$ if and only if $P_1,\ldots, P_n \vdash_D Q$. In such a case the difference between entailment Autonomy and derivability Autonomy (relative to D) makes no difference. A counterexample to entailment Autonomy is a counterexample to derivability Autonomy and vice versa.

Completeness is, however, a tall order. Derivation systems are at best complete relative to some limited class of entailments—for example, entailments that are valid solely in virtue of the nature of the truth functions, or in virtue of those together with the quantifiers. And there are barriers to completeness as soon as we enter the realm of simple arithmetic. Fortunately, for our purposes, these barriers do not matter. Suppose we discover that a normative proposition O is derivable from natural proposition N, within some sound system of derivation D: $N \vdash_D O$. Given D's soundness, it follows that $N \vDash O$ is valid, and we have a counterexample to entailment Autonomy. Suppose we discover that some natural proposition N entails a non-trivial normative proposition O. Then we of course have a counterexample to entailment Autonomy. However, we may not yet have a counterexample to derivability Autonomy relative to some system D. For it may be $N \nvdash_D O$, despite the validity of $N \vDash O$. The lack of derivability of $N \vDash O$ in D is clearly a defect of D. $N \vDash O$ would be derivable in a more adequate derivational system and whatever system of derivation we start with, we can augment it with rules to facilitate the derivation of $N \vDash O$. So throughout we will focus on entailment Autonomy.

3 Prior's Gem

Prior starts with a pair of propositions, **N** and **O**, where **N** is an indisputably natural proposition, and **O** is an indisputably normative proposition.

N: *All New Zealanders drink tea*.
O: *All New Zealanders ought to drink tea*.[9]

$\mathbf{N} \vDash \mathbf{O}$ is clearly not a valid entailment. It is possible that all New Zealanders drink tea even though not all ought to. A valid entailment featuring **N** amongst the premises and **O** as conclusion would need at least one additional premise, and that premise would have to contain some information about what ought to be done or what ought to be the case.

[6] Classes can be identified with mappings from a domain to the truth values. So this is tantamount to the thesis that the content of a proposition is a class of world-times. I include partial mappings here, something that is not always embraced by possible-worlds accounts of propositions. If a proposition induces a partial mapping from world-times to truth values, at those world-times at which it is undefined the proposition is truthvalueless.

[7] It will be convenient to also talk of entailment between the contents of propositions: $P^{\mathbf{CON}}$ entails $Q^{\mathbf{CON}}$ if and only if P entails Q. Since mappings from world-times to truth values are just (possibly partial) classes, this is also tantamount to: $P^{\mathbf{CON}} \subseteq Q^{\mathbf{CON}}$.

[8] Admittedly it is common for mathematicians to say things like: "the function $\lambda y.exp(y + 1{,}2)$ contains as constituents both the addition and exponentiation functions". But here they are not really talking about a *function*—a mapping from pairs of numbers to numbers. Rather, they are talking about a certain way of getting at, or "constructing", a function via some other functions. A different way of arriving at the very same function is this: $\lambda y.(y\times(y + 2) + 1)$. This latter construction of the function does contain as constituents the addition and multiplication functions, but does not contain exponentiation. See (Tichý 1986) for an informal exposition of the distinction between functions and constructions, and (Tichý 1988) for an extended analysis within transparent intensional logic.

[9] My example is rather gentler than Prior's. He uses: *all New Zealanders ought to be shot*.

For example, one might add the (somewhat implausible) premise that *New Zealanders only ever do what they ought to*. However, the following *is* valid, at least according to most logicians

All New Zealanders drink tea.

Therefore

Either all New Zealanders drink tea or they ought to.

Call this *Tea drinker* 1 ($\mathbf{T}_1$ for short): $\mathbf{N} \vDash \mathbf{N} \vee \mathbf{O}$.

$\mathbf{T}_1$ is an instance of an entailment schema taught with great confidence at the beginning of every introductory logic course.[10]

There are two interesting ways to challenge the validity of $\mathbf{T}_1$. One is based on relevantist concerns, and the other on the possibility of truth value gaps. I will put the former aside for the purposes of the discussion here, although it opens up an interesting alternative analysis of Autonomy and naturalism.[11] Suppose that you are an error theorist, or nihilist, about normative matters. You take normative sentences to have propositional content, and those propositional contents can happily serve both as the objects of thought and as the meanings of normative sentences. However, you think that no moral propositions are true. According to Mackie's early error theory, all substantive normative propositions are false.[12] But that's problematic. The negation of a normative claim (e.g. *not all New Zealanders ought to drink tea*) also seems to be a normative claim. So the negation of a normative falsehood seems to be a normative truth. Later Mackie claimed that no substantive normative claims are true.[13] The two positions are equivalent only if there are no truth value gaps. A better way for the nihilist to go here is to say that substantive normative propositions *lack* truth values, and they lack them because they suffer from presuppositional failure.[14] P is a presupposition of Q if the truth of P is necessary for Q to have a truth value. If negation is a truth function, then the negation of a proposition that lacks a truth value also lacks a truth value.[15]

In any case, there are good arguments independent of the demands of nihilism to embrace semantic gaps, of which truth value gaps are just one example. Consider the question: *What is the number of hairs on the head of the King of France?* Clearly the correct answer cannot be 0, for then it would follow that the King of France is bald, and that is not true. In fact there is no number that satisfies this description. The magnitude *the number of hairs on the King of France's head* does not yield any number as value in the actual circumstances. Now consider the answer to the question: *Is the number of hairs on the King of France's head even*? The correct answer is clearly not *true*. If it were *false* then, since every natural number is either even or odd, *the number of hairs on the King of France's head is odd* would presumably be true. But that is not true either. Suppose both are false. Then we would be able to infer that *there is a natural number—viz. the number of hairs on the King of France's head—that is neither odd nor even* contrary to an elementary number-theoretic fact.

The best solution to this puzzle is, I submit, to allow gaps, including truth value gaps.[16] *The King of France* and *the number of hairs on the King of France's head* have different extensions at different world times. But at some world-times they have no extensions at all. They induce partial functions from world-times to extensions. Likewise, the proposition—*the number of hairs on the King of France's head is even*—has as its extension (if it has an extension) one of the two truth values. But at those world-times at which *the number of hairs on the King of France's head* does not yield an extension, the proposition also has no extension. It yields no truth value.

Suppose **Ought** is a property of propositions.[17] A proposition P ought to be the case at some world-time just in case P is in the extension of **Ought** at that world-time. A nihilist could claim that **Ought** fails to have an extension at certain world-times. At such world-times its extension is not the empty class of propositions. *It has no extension at all*. **O** is just the proposition **Ought**(**N**), and would fail to yield a truth value at nihilistic world-times, as would its negation. Now, since disjunction is a truth function, the disjunctive proposition **N**∨**Ought**(**N**) fails to yield a truth value at any world-time at which **Ought**(**N**) fails to yield a truth

[10] An entailment schema is valid if every instance of it is valid. Even if the entailment schema is invalid this instance of it might be valid.

[11] See (Mares 2010) for a relevantist construal of Autonomy that renders it compatible with global supervenience, and hence with a version of naturalism.

[12] Mackie (1946).

[13] Mackie (1977).

[14] See (Oddie 2005) Chap. 1.

[15] Negation takes truth values to truth values. As King Lear noted, nothing comes from nothing. One cannot apply a function like negation to nothing at all and expect to end up with a truth value.

[16] Russell's theory, of course, offers a different solution to the puzzle, and it does entail that *the number of hairs on the head of the King of France is neither even nor odd*. That this sounds odd is, I think, a defect of the theory but plenty have been able to live with it. However, it does have other more serious defects. Russell's theory imputes existential import to propositions involving descriptions when the description occurs in a subordinate clause that is not propositional. That I am meditating on the number of hairs on the King of France's head does not entail that the King of France, or the number of hairs on his head, exists. But on Russell's account it does. The problem is ineradicable because not all attitudes can be parsed as propositional.

[17] *Ought* is more plausibly a property of certain properties, but I go here with the flow in deontic logic.

value—whatever the truth value of **N** at that world-time.[18] A nihilist could thus happily deny that $\mathbf{T}_1$ is valid. Whenever the first disjunct is true and the second is truth valueless (as it will be at nihilistic world-times) the premise is true and the conclusion is not. So if a normative proposition like **O** is truthvalueless at some world-time, $\mathbf{T}_1$ is not valid.

There are ways around this objection. One can close the truthvalue gap, while preserving the shape of the argument, by taking the second disjunct to be: *it is true that all New Zealanders ought to drink tea* (**True**(**O**)). If P is true at a world-time, the proposition, **True**(P) is also true at that world-time. But at any world-time at which P is false or truthvalueless **True**(P) is false. Further, if P is a normative proposition then **True**(P) is also normative, and so we can proceed throughout with the everywhere defined proposition **True**(**O**), instead of **O**.[19]

Back to Prior's gem. What is the natural/normative status of **N**∨**O**? If **N**∨**O** is normative then $\mathbf{T}_1$ is a violation of Autonomy—the natural proposition **N** entails the normative proposition **N**∨**O**. But suppose that **N**∨**O** is not normative. Consider *Tea drinker* 2:

($\mathbf{T}_2$): ¬**N**, **N**∨**O** ⊨ **O**.

$\mathbf{T}_2$ is valid.[20] Grant that if a proposition is not normative then it is natural. So **N**∨**O** is a natural proposition. Suppose further (as seems plausible) that the negation of a natural proposition is a natural proposition. Then ¬**N** is also natural. So in $\mathbf{T}_2$ we have two natural premises and a normative conclusion. On the second horn, $\mathbf{T}_2$ is a violation of Autonomy.

On whichever side of the natural-normative divide we opt to place **N**∨**O**, we seem to have at least one violation of Autonomy.

Many find Prior's dilemma unconvincing and yet saying what is wrong with it has proved quite hard. A number of commentators have constructed rather sophisticated proofs to show that, once we properly characterize what's going on, neither of these entailments breaks through the natural-normative entailment barrier.[21] Most deem $\mathbf{T}_1$ and $\mathbf{T}_2$ valid, but deny that they involve natural premises and a normative conclusion.

Inevitably, even as Prior's original inferences have been deemed wanting others have sprouted up to take their place. Consider the following rather obvious apparent counterexample (call it *Dupont*)[22]

Dupont didn't kill Schultz.

Therefore

Dupont didn't murder Schultz.

Killing is a purely natural relation so, unlike $\mathbf{T}_2$, the premise of *Dupont* is indisputably natural. Murdering is a particular kind of killing—a killing that you ought not to carry out. So unlike $\mathbf{T}_1$ the conclusion of *Dupont* does seem to make a non-trivial normative claim.

Believing, we can assume, is also a natural relation, one that holds between an individual and a proposition. A proposition about which beliefs Aunt Dahlia entertains is also presumably natural. Now consider the following (call it *Dahlia* [23]):

Aunt Dahlia believes that all New Zealanders ought to drink tea.

Everything Aunt Dahlia believes is true.

Therefore

All New Zealanders ought to drink tea.

By the principle of the naturalness of belief, the first premise is a natural proposition. Now consider the negation of the second premise—the proposition that *not everything that Aunt Dahlia believes is true*. A naturalist might happily affirm that with confidence, without tacitly rejecting naturalism. If the second premise is the negation of a proposition that naturalists can happily accept then it seems

[18] Since ∨ is a truth function and P and Q are not truth values but propositions, the logical form of the disjunction of P and Q is not perspicuously represented as $P \vee Q$ or $\vee PQ$. Where P is a construction of a mapping f from world-times to truth values, let $[P_{wt}]$ be the application of f to the pair w,t. If f is defined at w,t then $[P_{wt}]$ is a construction of the value of f at w,t. Let $[\vee P_{wt}\ Q_{wt}]$ be the application of the disjunction *function* ∨ to whatever pair of truth values (if there is such a pair) constructed by $[P_{wt}]$ and $[Q_{wt}]$. Then $\lambda w\ \lambda t\ [\vee P_{wt}\ Q_{wt}]$ gives the logical form of the disjunction of P and Q. $\lambda w\ \lambda t\ [\vee P_{wt}\ Q_{wt}]$ constructs a function g that takes w,t to true whenever both P and Q have truth values at w,t and one of those is true; false whenever both P and Q have truth values at w,t and either of those is false; and is undefined whenever either P or Q fails to yield a truth value at w,t.

[19] Note that if equivalence is taken to be mutual entailment then even with truth value gaps P and **True**(P) are equivalent since the following two entailment schemas are valid: P ⊨ **True**(P) and **True**(P) ⊨ P. Despite this, P and **True**(P) may induce distinct mappings from world-times to truth values, since **True**(***P***) is false whenever ***P*** is truthvalueless. So the following are not valid entailment schemas: ¬**True**(***P***) ⊨ ¬***P***; ¬**True**(***P***) ⊨ **True**(¬***P***). In the first case the premise can be true while the conclusion is truthvalueless. In the second case, the premise can be true while the conclusion is false.

[20] If **O** is truthvalueless so is **N**∨**O**. Whenever both premises are true so is the conclusion.

[21] See, for example, the papers in Pigden (2010) by Pigden, Schurz, Mares, and Restall and Russell.

[22] I cannot now locate the source of such examples. Perhaps it was the oral tradition.

[23] Adapted from Nelson (1995).

to follow that the second premise is also natural. If that is right *Dahlia* is another counterexample to Autonomy.

4 The Natural and the Normative

In order to determine the status of putative counterexamples to Autonomy we need an analysis of the concepts of the natural and the normative. Much of the work in defusing Prior's dilemma has been devoted to this task, but none appears to have garnered widespread acceptance. My analysis is somewhat novel but it is based on a more general account of subject matters.

Prior tacitly assumes what is now called *taxonomic essentialism*— that the natural/normative status of a proposition does not vary with the facts, and it does not vary from one context to another. If $\mathbf{N}\vee\mathbf{O}$ is a natural (or normative) proposition then it is natural (or normative) in every world and in every context, including in the contexts of different entailments, like $\mathbf{T}_1$ and $\mathbf{T}_2$. Taxonomic essentialism has been rejected by some Autonomists. Pigden and Schurz argue that the natural-normative status of a proposition must change with context on pain of contradiction.[24] According to Pigden, for example, $\mathbf{N}\vee\mathbf{O}$ is *natural* when it occurs in the conclusion of $\mathbf{T}_1$—hence both premise and conclusion are natural, so there is no violation of Autonomy. But it is *normative* when it occurs in the premises of $\mathbf{T}_2$. Hence the premises of $\mathbf{T}_2$ include a normative proposition, and again we have no violation of Autonomy.

While there are interesting and subtle arguments for this contextualism, it strikes me as a little counterintuitive.[25] The account I will give of the natural-normative distinction yields taxonomic essentialism. This not only has the advantage of being intuitively more natural, but it provides an initially more charitable gloss on Prior's argument.

A closely related assumption is that the natural-normative distinction is an intensional one, not a hyperintensional one. That is to say: if P is a natural (respectively: a normative) proposition then any proposition Q with the same content as P is also a natural (respectively: normative) proposition. Whether a proposition is normative or natural depends on what possibilities it rules in or out, not on the particular way it goes about doing so.

Consider the following inference, *Tea drinker* 3:

$(\mathbf{T}_3)$: $\mathbf{N} \vDash \mathbf{N}\vee\neg\mathbf{N}$.

$\mathbf{T}_3$ is valid, obviously, even if we allow truth value gaps. Further, both premise and conclusion seem natural. Now consider *Tea drinker* 4:

$(\mathbf{T}_4)$: $\mathbf{N} \vDash \mathbf{O}\vee\neg\mathbf{O}$.

The conclusion, $\mathbf{O}\vee\neg\mathbf{O}$ seems as normative as $\mathbf{N}\vee\neg\mathbf{N}$ seems natural, but if $\mathbf{O}\vee\neg\mathbf{O}$, were a normative proposition $\mathbf{T}_4$ would violate Autonomy. Something is clearly wrong here, and the intensional constraint puts a finger on part of it, but not all. $\mathbf{O}\vee\neg\mathbf{O}$ is standardly taken to be equivalent to $\mathbf{N}\vee\neg\mathbf{N}$, and so, by the intensional constraint, $\mathbf{O}\vee\neg\mathbf{O}$ is normative if and only if $\mathbf{N}\vee\neg\mathbf{N}$ is normative.[26] If $\mathbf{N}\vee\neg\mathbf{N}$ is a normative proposition then $\mathbf{T}_3$ is already a counterexample to Autonomy and that would be silly. However, if $\mathbf{N}\vee\neg\mathbf{N}$ is a natural proposition then so too is $\mathbf{O}\vee\neg\mathbf{O}$, and $\mathbf{T}_4$ would not be violation of Autonomy after all.

But if $\mathbf{N}\vee\neg\mathbf{N}$ is natural, and the negation of a natural proposition is natural, (as Prior seems to assume in the second horn of his dilemma), then $\neg(\mathbf{N}\vee\neg\mathbf{N})$ is also natural, and (again by the intensional constraint) so too is $(\mathbf{N}\wedge\neg\mathbf{N})$. We would then have the following very cheap violations of Autonomy:

$(\mathbf{T}_5)$: $\neg(\mathbf{N}\vee\neg\mathbf{N}) \vDash \mathbf{O}$; $(\mathbf{N}\wedge\neg\mathbf{N}) \vDash \mathbf{O}$.

And even without Boolean closure, we have the closely related valid entailment:

$(\mathbf{T}_6)$: $\mathbf{N}, \neg\mathbf{N} \vDash \mathbf{O}$.

$\mathbf{T}_5$ and $\mathbf{T}_6$ raise two issues—one concerning the natural/normative status of necessarily false propositions, and the other concerning the status of entailments with incompatible premises.

Prior appears to assume Boolean closure of the natural and the normative. And Boolean closure would seem like a plausible requirement for the non-naturalist to embrace anyway.[27] If you start with some natural (or normative) propositions you cannot break out of the natural (or normative) realm by negating, disjoining and conjoining what you start with. So, for example, if $\mathbf{N}$ is a natural proposition then so too are $\neg\mathbf{N}$, $\mathbf{N}\vee\neg\mathbf{N}$ and $\mathbf{N}\wedge\neg\mathbf{N}$. More generally:

> *Boolean Closure*.
> If P and Q are natural (normative) propositions then $\neg P$, $P\vee Q$ and $P\wedge Q$ are also natural (normative).

Given this, necessarily true and necessarily false propositions are both natural and normative. But necessarily false

[24] Pigden 1989 and Schurz 1997. For a summary of their contributions see (Pigden 2010), pp 33–36.

[25] For doubts about this contextualist strategy see (Brown 2014).

[26] This would not be the case if there are truth value gaps. For then $\mathbf{O}$, $\neg\mathbf{O}$ and $\mathbf{O}\vee\neg\mathbf{O}$ might all fail to have a truth value at some world-time at which $\mathbf{N}$ is true. Again, however, we could replace $\mathbf{O}$ with $\mathbf{True}(\mathbf{O})$ and $\mathbf{N}$ with $\mathbf{True}(\mathbf{N})$. $\mathbf{True}(\mathbf{O})$ is normative, $\mathbf{True}(\mathbf{N})$ is natural, and $\mathbf{True}(\mathbf{O})\vee\neg\mathbf{True}(\mathbf{O})$ is equivalent to $\mathbf{True}(\mathbf{N})\vee\neg\mathbf{True}(\mathbf{N})$.

[27] For an argument against the Boolean closure of the class of natural *properties* see (Oddie 2005), Chap. 6. If natural properties carve out convex regions of the natural space, then, since negation and disjunction do not preserve convexity, natural properties are not closed under these operations. This is a very strong notion of naturalness—too strong, I think, to capture a reasonable account of Autonomy and Entanglement.

propositions and incompatible premises entail every proposition. The Autonomist can fend off the cheap counterexamples with one of two strategies: either restrict Boolean Closure to the non-trivial cases (excluding necessarily true and necessarily false propositions from the scope of the closure clause); or restrict Autonomy, excluding the degenerate entailments that contain either necessarily false premise sets, or necessarily true conclusions. Even if we restrict Boolean closure, we are still going to have to restrict Autonomy to neutralize $\mathbf{T}_6$. Both premises of $\mathbf{T}_6$ are natural and neither is necessarily false. Of the two options I propose we accept Boolean closure (for simplicity) and refine Autonomy accordingly.

Note that if we accept Boolean closure, and the intensional constraint, then all necessarily true propositions and all necessarily false propositions will be deemed both natural and normative—since there is at least one necessarily true natural proposition and one necessarily false normative proposition. This may seem artificial, even counterintuitive, but the cost is negligible, since the refinement of Autonomy sidelines them.

Call a collection of propositions *consistent* if they are all true at some world-time, and call a proposition *trivial* if it is necessarily true. Then Autonomy can be refined thus:

> *Autonomy*.
> No consistent collection of natural (respectively: normative) premises entails a non-trivial normative (respectively: natural) proposition.

$\mathbf{T}_4$ has a natural premise and a normative conclusion, but its conclusion is trivial. The premises of $\mathbf{T}_5$ and $\mathbf{T}_6$ are also natural (given Boolean closure), but in neither case are they consistent. The cheap counterexamples do not refute this more refined formulation of Autonomy. But sadly we are no closer to a resolution of Prior's dilemma. If we grant taxonomic essentialism, as I have, then if $\mathbf{N}\vee\mathbf{O}$ is normative, $\mathbf{T}_1$ is a violation of Autonomy, and if $\mathbf{N}\vee\mathbf{O}$ is natural, $\mathbf{T}_2$ is a violation.

5 Questions, Answers and Subject Matters

What we clearly need is an account of the natural and the normative which settles the issue of the natural/normative status of the particular propositions at issue.

Assume that we have a space of maximal possibilities, or world-times, relativized to certain parameters—like a domain of fundamental entities and some fundamental properties and relations of such entities. (One could instead start with situations, or parts of worlds, but I will not pursue that here.) Each space of possibilities generates various *questions* or *subject matters*. Questions like: *What's the number of the planets? What's the weather in Boulder? What are the laws of motion?* These questions, we can suppose, admit of a range of possible answers. More generally, a class T of traits (properties, relations, magnitudes and so on) defines a *s*ubject matter and an associated *T-question*, namely: *What is the actual distribution or extensions of the T-traits?*[28]

If there is no way of drawing the distinction between normative and natural traits then of course Autonomy is doomed from the get-go. Typically the distinction is drawn at the syntactic level, by assuming a division of *predicates* into two classes—the clearly natural predicates (like "tea" and "drink" and "kill"), and a set of (thin) normative predicates (like "permissible", "wrong" and "good"). In addition there may be a set of mixed predicates (like "murder") along with the rich vocabulary that we employ for the so-called *thick* value attributes. We evaluate people as *courageous, compassionate, callous, cruel, charming* and *sexy*. We evaluate actions as *generous, vindictive, kind* and *foolhardy*. We evaluate performances as *brilliant, elegant, clumsy, riveting, delightful* and *poised*. We evaluate remarks as *tendentious, salacious, witty, craven, hurtful, sarcastic, biting* and *helpful*. That there is this rich set of predicates denoting attributes which span the natural-normative divide has been cited as a reason to eschew non-naturalism and embrace logical leakage, denying the existence of a strict natural-normative divide. As we will see, this may well be a reason to eschew an extreme non-naturalism, one that goes hand-in-glove with Autonomy. And it does so by undermining the thesis that there is a clean natural-normative divide at the level of properties, which is necessary if there is to be a clean natural-normative divide at the level of propositions.

To what class should apparently mixed predicates, and the properties they denote or express, be assigned? Autonomists typically think that mixed predicates decompose neatly into their more basic purely normative and poorly natural ingredients. "Murder", to take one of the simpler examples, might be analyzed as "impermissible killing of a person". If such analyses can be supplied then well and good, but if there is some deep reason they cannot be supplied then that will be some sort of count against the plausibility of Autonomy.

In order to give Autonomy a charitable hearing, I will assume that we do indeed have a domain of *basic* natural traits and, completely disjoint from that, a domain of *basic* normative traits. Further, we need to assume that

[28] The notion of subject matters as partitions was introduced in Oddie (1986) (108–111) and independently by David Lewis in his (1988), which in turn draws on the earlier work on questions by Belnap and Steele (1976). My analysis is not dissimilar from the one given in Brown (2014). But Brown argues for Autonomy and in the end I will show that it must fail, in any sufficiently rich framework.

any apparently mixed traits, those that do not fall clearly into either of these categories, can somehow be cashed out (reduce to or supervene upon) the union of these two sets of basic traits. Again, if this assumption fails then Autonomy is already in trouble. In order not to beg the question against the Autonomist let's assume the separability and sufficiency of the basic natural and normative traits.

The two questions, or subject matters, that we are interested in here are these: What is the natural structure of the world? (or: What are the extensions of the natural traits?); and: What is the normative structure of the world? (or: What are the extensions of the normative traits?) We can call the former the Natural Question and the latter the Normative Question.

Consider a very simple subject matter: *the number of the planets*. Where φ is the magnitude *the number of the planets*, and Q is the associated question, let φ_n be the class of all worlds in which the magnitude takes the value n. φ_n constitutes a complete answer to the question Q, and these complete answers are mutually exclusive and jointly exhaustive.[29] The question Q can be identified with the following partition of the logical space:

$$Q^{\varphi} = < \varphi = 0, \varphi = 1, \varphi = 2, \dots, \varphi = n, \dots > .$$

In general each subject matter Q can be identified with a partition $< C_0, C_1, \ldots >$ of the logical space, each cell C_i of which corresponds to one of the complete possible answers to the question Q. In each world-time of a single cell C_i of the partition, the true answer to the question is just C_i.

The Natural Question (**NAT**) partitions the logical space into cells $< N_0, N_1, \ldots, N_n, \ldots >$ each member of which contains world-times that share exactly the same natural structure. In any two elements of N_i all the natural attributes have exactly the same extensions. Similarly the Normative Question (**NORM**) partitions the logical space into cells $< O_0, O_1, \ldots, O_n, \ldots >$ the members of which contain world-times which share exactly the same normative structure. In any two world-times in O_i the normative attributes will have exactly the same extensions.

Each cell of a question Q provides a complete (correct or incorrect) answer to the question Q. From the point of view of Q, the world-times that share a cell are indistinguishable, they yield the same answer to the question under consideration. But not all answers to a question need be complete. If you ask what is the number of the planets one possible response is: *it is either eight or nine;* another is *no more than ten*. These are quite good answers. They are both true as of now but they are partial. Some false complete answers (e.g. *the number of planets is nine*) also are not bad, as are some false partial answers (e.g. *the number of planets is either seven or nine*). A partial answer to Q carves out a set of complete answers—all those it does not rule out. Two propositions give the same answer to Q just in case they carve out the same subset of Q. A complete answer carves out a singleton subset, a partial answer carves out a non-empty subset. There are two limiting subsets of Q: the set of all complete answers (this is the trivially true answer to Q), and the empty set of complete answers (this is the trivially false answer to Q). The former doesn't rule out any complete answer. It is maximally partial, minimally complete. It is the minimally informative answer. The latter rules out all complete answers, it is neither a complete answer nor a partial answer. Are these two limiting cases any sort of answer at all? There is a certainly case for answering in the negative here, ruling them both out as degenerate. I include them largely for simplicity.

For each proposition P and question Q, P is compatible with certain Q-cells and incompatible with the rest. *P's answer to Q*—what I will also call P's *Q-content*—is the union of all the Q-cells that P leaves open as live contenders for the complete answer to Q. Every proposition has an answer (possibly a degenerate answer) to each question. Two propositions give the same answer to Q just in case they leave open the same range of complete answers to Q. Thus two logically equivalent propositions yield the same answers to all questions.

Recall that the content of proposition P—$P^{\mathbf{CON}}$—is a set of world-times—or, what is the same thing, a function from world-times to truth values. P's Q-content is also a function from world-times to truth values and is, intuitively, the content of *P's* answer to Q. P's *natural* content is the content of its answer to the natural question. $P^{\mathbf{NAT}}$ takes a world-time to true if it lies in a cell N_i compatible with P, and false if it lies outside all such cells. Similarly, *P's normative* content takes a world-time to true if it lies within a cell O_i of the normative partition compatible with P, and to false if it lies outside all such cells. Note that, if a proposition's content assigns true to a world-time, then so too does its Q-content, though the converse does not always hold. The content of P is always as strong as its Q-content, for any Q, and may well be stronger.

We can now define the class of *natural* propositions. Suppose a proposition's content exceeds its natural content. Such a proposition says more than just what it says about the natural. Such a proposition is clearly not (purely) natural. But if a proposition says no more than what it says about the natural (and it cannot say less) then that is

[29] Well, almost. There may be no number n such that n numbers the planets in the Solar System —whenever the Solar System does not exist (e.g. one second after the Big Bang). It is not that there are 0 planets in the Solar System at that world-time, but that the question doesn't even arise at that world-time. It contains a false presupposition. So we need a further element of the partition, one which corresponds to the answer: *the question does not arise* or *does not have a value* at the world-time in question.

a sufficient condition for its being a natural proposition. A proposition is thus natural just in case its natural content coincides with its content. (*Mutatis mutandis* for the *normative*.) In general, a Q-proposition (a proposition wholly about the subject matter Q) is any proposition whose content and Q-content coincide.

Natural and normative.
P is a *natural* proposition if and only if $P^{\mathbf{CON}} = P^{\mathbf{NAT}}$.
P is a *normative* proposition if and only if $P^{\mathbf{CON}} = P^{\mathbf{NORM}}$.

This account yields taxonomic essentialism. Firstly, the status of a proposition is world-time independent, and context-independent. Secondly, if two propositions have the same content then they have the same Q-content for any Q. Hence *natural* and *normative* are non-hyperintensional, world-independent and non-contextual, as desired. This account also delivers the Boolean closure of the class of Q-propositions. If P is a natural proposition then its content coincides with its natural content. P maps a world-time in cell N_i to true (false) just in case it maps all world-times in cell N_i to true (false). The content of $\neg P$ reverses P's mapping. $\neg P$ maps a world-time in cell N_i to false (true) just in case it maps all world-times in cell N_i to false (true). So the content of $\neg P$ also coincides with its natural content. We can show similarly that if P and Q are both natural so too are $P \vee Q$ and $P \wedge Q$.

6 Prior's Dilemma Analyzed

We have two partitions induced by the natural and the normative questions. Every proposition P has both a natural content $P^{\mathbf{NAT}}$ and a normative content $P^{\mathbf{NORM}}$. P is a natural proposition just in case P's content coincides with its natural content $P^{\mathbf{NAT}}$ (P's content does not breach the divides of the natural partition), and it is a normative proposition just in case P's content coincides with its normative content $P^{\mathbf{NORM}}$ (P does not breach the divides of the normative partition).

We will work with a very simplified model. Let's suppose both partitions contain just two cells apiece: $\mathbf{NAT} = \{\mathbf{N}, \neg\mathbf{N}\}$; $\mathbf{NORM} = \{\mathbf{O}, \neg\mathbf{O}\}$, and suppose further that all the cells in the product partition $\{\mathbf{N}\wedge\mathbf{O}, \mathbf{N}\wedge\neg\mathbf{O}, \neg\mathbf{N}\wedge\mathbf{O}, \neg\mathbf{N}\wedge\neg\mathbf{O}\}$ are non-empty. Then $\mathbf{N}$, $\neg\mathbf{N}$, $\mathbf{N}\vee\neg\mathbf{N}$ and $\mathbf{N}\wedge\neg\mathbf{N}$ (and their logical equivalents) are the natural propositions, while $\mathbf{O}$, $\neg\mathbf{O}$, $\mathbf{O}\vee\neg\mathbf{O}$ and $\mathbf{O}\wedge\neg\mathbf{O}$ (and their logical equivalents) are the normative propositions. The natural/normative classification is clearly neither exclusive nor exhaustive. Necessarily true and necessarily false propositions are both natural and normative, as one would expect given Boolean closure. Some propositions, like $\mathbf{N}\wedge\mathbf{O}$ and $\mathbf{N}\vee\mathbf{O}$, feature in neither category. $\mathbf{N}\wedge\mathbf{O}$ is the conjunction of a natural proposition $\mathbf{N}$ and a normative proposition $\mathbf{O}$. $\mathbf{N}\wedge\mathbf{O}$ has non-trivial natural content ($=\mathbf{N}^{\mathbf{NAT}}$) and non-trivial normative content ($=\mathbf{O}^{\mathbf{NORM}}$). $[\mathbf{N}\wedge\mathbf{O}]^{\mathbf{CON}}$ coincides neither with $[\mathbf{N}\wedge\mathbf{O}]^{\mathbf{NAT}} = \mathbf{N}^{\mathbf{NAT}}$ nor with $[\mathbf{N}\wedge\mathbf{O}]^{\mathbf{NORM}} = \mathbf{O}^{\mathbf{NORM}}$. But the content of $\mathbf{N}\wedge\mathbf{O}$ does decompose neatly into its natural and normative components. Where f and g are mappings from world-times to truth values, the *fusion* of f and g ($f\otimes g$) is the function that takes a world-time to true if both f and g take it to true, is undefined if either of f or g is undefined, and takes it to false otherwise.[30] A proposition P that is neither natural nor normative but whose content is the fusion of its natural and normative contents we call a *natural-normative fusion*.

Fusion.
P is a *fusion* if and only if: $P^{\mathbf{CON}} \neq P^{\mathbf{NAT}}$, $P^{\mathbf{CON}} \neq P^{\mathbf{NORM}}$ and $P^{\mathbf{CON}} = P^{\mathbf{NAT}} \otimes P^{\mathbf{NORM}}$.

The content of some propositions does not decompose neatly into natural and normative components. $\mathbf{N}\vee\mathbf{O}$ entails no non-trivial natural proposition, and no non-trivial normative proposition. Its natural content and its normative content are both trivial. $[\mathbf{N}\vee\mathbf{O}]^{\mathbf{NAT}} = [\mathbf{N}\vee\mathbf{O}]^{\mathbf{NORM}} = [\mathbf{N}\vee\neg\mathbf{N}]^{\mathbf{CON}}$. Consequently their fusion— $[\mathbf{N}\vee\mathbf{O}]^{\mathbf{NAT}} \otimes [\mathbf{N}\vee\mathbf{O}]^{\mathbf{NORM}}$—is also trivial. $\mathbf{N}\vee\mathbf{O}$, however, has non-trivial content. $[\mathbf{N}\vee\mathbf{O}]^{\mathbf{CON}}$ maps all world-times in the cell (corresponding to) $\neg\mathbf{N}\wedge\neg\mathbf{O}$ to false. So $[\mathbf{N}\vee\mathbf{O}]^{\mathbf{CON}} \neq [\mathbf{N}\vee\mathbf{O}]^{\mathbf{NAT}} \otimes [\mathbf{N}\vee\mathbf{O}]^{\mathbf{NORM}}$. Such propositions are not fusions but *hybrids*.[31]

Hybrid.
P is a *natural-normative hybrid* if and only $P^{\mathbf{CON}} \neq P^{\mathbf{NAT}} \otimes P^{\mathbf{NORM}}$.

We now have a four-fold rather than two-fold classification of propositions. Natural propositions (those the content of which is identical to their natural content); normative propositions (those the content of which is identical to their normative content); natural-normative fusions (those the content of which is the fusion of non-trivial natural and normative contents); and natural-normative hybrids (those the content of which is not the fusion of natural and normative contents). This classification is exhaustive: every proposition is either natural, or normative, or a natural-normative fusion or a natural-normative hybrid. It is almost exclusive, except for the limiting cases of necessarily true and necessarily false propositions, which are both natural and normative.

[30] To preserve the appropriate content for propositions with truth value gaps, for any contents C and D, $C\otimes D$ must be *undefined* at any world-time at which either C or D is undefined.

[31] The distinction between fusions and hybrids was introduced in Oddie and Demetriou (2007).

We can use the classification to diagnose what exactly is going on in Prior's dilemma. Prior tacitly assumes that every proposition must be classified as either natural or normative but not both. Given that **N** and ¬**N** are natural and **O** is normative, in order to avoid violating the naive Autonomy thesis, $\mathbf{T}_1$ requires that **N**∨**O** be classified as natural, while $\mathbf{T}_2$ requires that **N**∨**O** be classified as normative. On our more fine-grained analysis, however, **N**∨**O**, while non-trivial, is not a normative proposition. So $\mathbf{T}_1$ does not feature a conclusion that is both non-trivial and normative. Nor is **N**∨**O** a natural proposition, so while $\mathbf{T}_2$ has consistent premises, they are not all natural. Rather, **N**∨**O** is a natural-normative hybrid. Thus no violation of Autonomy can be wrung out of Prior's dilemma.

Since **N**∨**O** is neither a natural proposition nor a normative proposition, it is both non-natural and non-normative. The Autonomist might object that $\mathbf{T}_1$ violates the Autonomist intuition that a natural premise cannot entail a *non-natural* conclusion. And $\mathbf{T}_2$ violates the intuition that premises all of which are non-normative cannot entail a normative conclusion.[32] But these maxims only seem plausible on a coarse and inadequate classification of propositions. The correct classification exposes the mistake here. While a natural-normative hybrid is not purely natural, there is nothing at all odd in the fact that a purely natural proposition entails some natural-normative hybrids. Further there is nothing odd in the fact that a natural proposition together with a natural-normative hybrid (e.g. the natural proposition ¬**N** together with the hybrid proposition **N**∨**O**) entails a normative proposition (**O**).

N∧**O** is neither natural nor normative, despite the fact it has non-trivial natural and normative content. Wouldn't it be more accurate to say it is *both* natural and normative? At this stage, we can improve the terminology to better capture the underlying Autonomist intuition. Let us say that a proposition is *purely natural* if it is a natural proposition but it is not also a normative proposition (*mutatis mutandis* for the purely normative). Since necessarily true and necessarily false natural propositions are both natural and normative they are not purely natural or purely normative. So, every purely natural proposition is a proposition with non-trivial natural content and no non-trivial normative content. There are clearly propositions that, while not purely natural, nevertheless have non-trivial natural content. So, propositions with non-trivial natural content divide into those that are purely natural and those that are not purely natural. (Likewise for propositions that are not purely normative but have non-trivial normative content.) The latter are all either fusions or hybrids.

With this nomenclature we can reformulate Autonomy:

> *Autonomy*
> No collection of purely natural (normative) premises entails a purely normative (natural) proposition.

Given Boolean closure the conjunction of any collection of purely natural propositions is itself a purely natural proposition. So, if for every collection of propositions there is a proposition which is the conjunction of those, Autonomy is equivalent to the simpler:

No purely natural (normative) proposition entails a purely normative (natural) proposition.

Fusions and hybrids are neither purely natural propositions, nor purely normative propositions, and so it should come as no surprise to the Autonomist that these can figure in the premises of entailments that boast purely normative conclusions, or in the conclusions of entailments that boast purely natural premises.

Where does this leave Prior's gem? While not a refutation of Autonomy, $\mathbf{T}_1$ illustrates the fact that natural-normative hybrids are entailed by purely natural premises. But that is hardly shocking. What is interesting, however, is that there can be pairs of propositions, both of which are *empty* of any normative content, which jointly entail a non-trivial purely normative proposition. **N**∨**O** is empty of normative content. So too is ¬**N**. Nevertheless ¬**N** and **N**∨**O** jointly entail **O**. And we can strengthen this result. Consider these two valid entailments:

($\mathbf{T}_7$): **N∨O**, **¬N∨O** ⊨ **O**.

($\mathbf{T}_8$): **N∨O, N∨¬O** ⊨ **N**.

Both feature two hybrid premises, none of which has any non-trivial natural or normative content. $\mathbf{T}_7$ boasts a purely normative conclusion and $\mathbf{T}_8$ a purely natural conclusion.

What about Dupont and Dahlia?

7 Dupont and Dahlia

Dupont is easily defused. Let **K** be the proposition that Dupont killed Schulz; **M**, that Dupont murdered Schulz; **W**, that Dupont's killing of Schulz would be morally wrong. **M** is equivalent to **K**∧**W**, and so **M** ⊨ **K** is clearly a valid entailment. **K** and ¬**K** are purely natural, while **W**, ¬**W** are purely normative (or so we can assume). The content of **K**∧**W** is the fusion of its natural and normative components, so it is a fusion. That a fusion entails non-trivial natural and normative propositions is no violation of either the Autonomist's intuition or of its articulation in Autonomy. *Dupont* is just the converse entailment, ¬**K** ⊨ ¬**M**. ¬**M** is equivalent to ¬**K**∨¬**W**, which, like **N**∨**O**, is a hybrid, the natural and normative contents of which are

[32] Pigden makes this objection to a classification similar to the one I am proposing here, although it is not based on subject matters and does not distinguish fusions and hybrids.

both empty. *Dupont* thus involves a purely natural proposition, $\neg\mathbf{K}$, entailing a hybrid proposition, $\neg\mathbf{K}\vee\neg\mathbf{W}$. *Dupont* is similar to $\mathbf{T_1}$ and thus poses no threat to Autonomy.

Dahlia is not as simple. Let **Believes**(P) be: *Aunt Dahlia believes P*.

Dahlia: $\mathbf{Believes}(\mathbf{O})$, $(\forall P)(\mathbf{Believes}(P)\supset P) \vDash \mathbf{O}$.

Belief is a natural trait, and its extension is part of the answer to the natural question. It follows that **Believes**(P) is purely natural, for any P whatsoever. The first premise of *Dahlia* is purely natural. Is the second premise a purely natural proposition? It is a generalization (**G**) to the effect that all Aunt Dahlia's beliefs are true. Consider a simpler and more modest *Dahlia*, one with the same first premise and conclusion, but with a pared down second premise, just one instance of the generalization: if Aunt Dahlia believes **O** then her belief about **O** is true.

Simple Dahlia: $\mathbf{Believes}(\mathbf{O})$, $\mathbf{Believes}(\mathbf{O})\supset\mathbf{O} \vDash \mathbf{O}$.

Simple Dahlia goes straight to that instance of **G** that meshes with the first premise to yield the normative conclusion. As such, if *Dahlia* is a counterexample to Autonomy then *Simple Dahlia* presumably is too. For it is only Aunt Dahlia's belief concerning **O**, not any of her other beliefs, that plays the crucial role in yielding **O** from **G**.

$\mathbf{Believes}(\mathbf{O})\supset\mathbf{O}$ is tantamount to $\neg\mathbf{Believes}(\mathbf{O})\vee\mathbf{O}$, where $\neg\mathbf{Believes}(\mathbf{O})$ is purely natural and **O** is purely normative. Like $\mathbf{N}\vee\mathbf{O}$, $\neg\mathbf{Believes}(\mathbf{O})\vee\mathbf{O}$ is a hybrid. So, *Simple Dahlia* involves a purely natural premise and a hybrid premise jointly yielding a normative conclusion. It is thus like $\mathbf{T}_2$, and so poses no threat to Autonomy. If *Dahlia* is a counterexample only if *Simple Dahlia* is too, *Dahlia* is no counterexample to Autonomy. I have not proved that **G** isn't purely natural. But at this stage we have no good reason to think it is purely natural. (We will return to this below.) So far, then, *Dahlia* is not a clear counterexample to Autonomy.

The upshot so far is that the normative and the natural can be interestingly entangled in the *contents* of hybrid propositions. But since this is fully compatible with Autonomy we can label this kind of entanglement *shallow*. The question now is whether there is a deeper natural-normative entanglement. Is there any entanglement at the level of states of affairs?

8 A Proof of Deep Entanglement

Autonomy articulates a certain very strong relation between the natural and normative questions. Suppose we have two non-trivial questions (both admit of more than one possible answer). And suppose that no complete answer to the one rules out any of the (complete or incomplete) answers to the other. Then the two questions are Autonomous. The underlying idea of Autonomy is basically just that every complete answer to the natural question is consistent with every complete answer to the normative question. But this is entirely compatible with Prior's shallow entanglement as we have seen. We now demonstrate that in any system that is minimally adequate for the representation of moral ontology, deep entanglement ensues and Autonomy fails.

There are basic traits, like *goodness* and *moral permissibility*, that all sides of the Autonomy debate concede are paradigmatically normative. (I will use *goodness* for concreteness.) And there are basic traits, like *belief* and *desire*, that are generally, even if not universally, held to be paradigmatically natural. (I will use *desire*.) A complete answer to the normative question—which I will call a *normative structure*—will specify (*inter alia*) the extension of *goodness*. And a complete answer to the natural question will specify (*inter alia*) the extension of *desire*.

If desire turns out not to be a natural trait, then the naturalist will be on the back foot from the start. Many prominent naturalist theories of goodness start with desires as the basic natural building blocks. For example, a naturalist might hold that a natural state is *better* the greater the number of satisfied desires it contains. Or that a natural state is *good* just in case the ratio of satisfied to unsatisfied desires is positive. But of course naturalists can disagree coherently about this. They countenance a wide range of different candidate normative structures some of which are more plausible than others, and the normative realist holds that one of them, perhaps as yet unknown to us, is the correct answer.

So, both naturalist and non-naturalist realists typically countenance a range of normative structures which assign different extensions to the normative traits. In particular, a normative structure (a complete answer to the normative question) will, for each natural structure N_j, either assign N_j to the extension of **Good**, or will exclude N_j from the extension of **Good**. Different normative structures deem different kinds of natural structures **Good**. It will be useful here to introduce the notion of a *P-norm*.

> *P-norms*.
> O_i is a *P-norm* just in case O_i assigns to the extension of **Good** all and only those complete answers to the natural question, N_j, that entail natural proposition P.

For example, let **S** be the proposition that there are no unsatisfied desires. O_i counts as an **S**-norm if it entails that for any natural state N_j, N_j is **Good** if and only, in all N_j-worlds, the extension of *unsatisfied desire* is empty. Or let **S*** be the less demanding proposition that there are more satisfied desires than unsatisfied desires. O_i counts as an **S***-norm if it entails that for any natural state N_j, N_j is **Good** if and only if, in all N_j-worlds, the ratio of satisfied to total number of desires is greater than 1/2. I am assuming here that *desire* is a basic natural relation between individuals and the objects of desire (propositions say) and that an individual's desire is satisfied (in a possible circumstance) just in case the content of her desire obtains in that circumstance (the proposition is true).

Consider a particularly simple *P*-norm. Let's fix on a particular individual (Charles, say) and Charles's occurrent desires. Let **Desire**(Q) be the proposition that Charles occurrently desires *P*. Since **Desire** is a basic natural trait, **Desire**(Q) is a purely natural proposition for each Q. Now, in certain possible circumstances Charles is fixated on just one object of desire: there is one and only one Q such that Charles desires Q. Let $\mathbf{Desire}^{\mathbf{U}}(Q)$ be the proposition that Charles uniquely desires Q. If *desire* is a natural relation then $\mathbf{Desire}^{\mathbf{U}}(Q)$ is a purely natural proposition. For if the extension of *desire* is the same in two possible circumstances, then $\mathbf{Desire}^{\mathbf{U}}(Q)$ must have the same truth value in both.

Let **Sated** be the proposition: *more of Charles's occurrent desires are satisfied than not*, and let O_i be any **Sated**-Norm. O_i entails that all and only those complete natural propositions (cells of the natural partition) are **Good** that entail that more of Charles's occurrent desires are satisfied than not. This is an outlandish normative theory, but some normative theories are not totally unlike this. Some do designate a particular individual whose desires play a special role in making the world a good one (for example, *God* or *Me*). In any case, while other normative structures could serve just as well, this one is nice and simple. Now let **NormSated** be a proposition that says that the normative structure of the world is a **Sated**-Norm. **NormSated** is clearly a normative proposition. If two possible circumstances have the same normative structure then **NormSated** has the same truth value in both.

Let **NotGood** be the proposition that *it is not true that the actual natural state is good*. As noted, for any Q, $\mathbf{Desire}^{\mathbf{U}}(Q)$ is purely natural. In particular, $\mathbf{Desire}^{\mathbf{U}}$(**NotGood**) is purely natural. Since **NormSated** is purely normative, so is its negation, ¬**NormSated**. We can now demonstrate the following:[33]

Theorem

$\mathbf{Desire}^{\mathbf{U}}$(**NotGood**) ⊨ ¬**NormSated**.

Proof

Assume (for the sake of a reductio) that $\mathbf{Desire}^{\mathbf{U}}$(**NotGood**) and **NormSated** are both true at world *W*. Notice that even at worlds at which the question of the truth value of the goodness of the natural state at that world does not arise (because, say, the normative structure of that world is a nihilist one) **NotGood** is nevertheless *true*. This is because any proposition of the form *it is not true that P* is true whenever *P* is either false or truthvalueless.

i. Suppose **NotGood** is *true* at *W*: it is either false or truthvalueless that the natural state of *W* is in the extension of **Good** at *W*. So, Charles's sole desire in *W* (for **NotGood**) is satisfied in *W*. Consequently, **Sated** is true at *W*. Given that **NormSated** and **Sated** are true in *W*, the natural state of *W* is in the extension of **Good** in *W*. So ***NotGood*** is false in *W*. (Contradiction.)
ii. Suppose **NotGood** is *false* in *W*: the natural state of *W* is in the extension of **Good** at *W*. Then one of Charles's occurrent desires in *W* (namely for **NotGood**) is *not* satisfied in *W*. Since Charles only has that one desire in *W*, **Sated** is false. Hence ¬**Sated** is true at *W*. Given **NormSated** and ¬**Sated** are both true in *W*, it follows that the natural state of *W* is not in the extension of **Good** in *W*. So **NotGood** is true in *W*. (Contradiction.)

9 Some Objections

Objection: If $\mathbf{Desire}^{\mathbf{U}}(Q)$ is not purely natural we do not have a violation of Autonomy. $\mathbf{Desire}^{\mathbf{U}}(Q)$ is purely natural if *and only if* the *object* of the desire, Q itself, is purely natural. Since **NotGood**, like **Good**, is purely normative, $\mathbf{Desire}^{\mathbf{U}}$(**NotGood**) is not purely natural.

Reply: As shown above, if *desire* is a basic natural trait then $\mathbf{Desire}^{\mathbf{U}}(Q)$ is a purely natural proposition regardless of the natural/normative status of Q. To sustain this first response the Autonomist would have to abandon the thesis that desire is a basic natural trait. But that is a very unstable position for the Autonomist to occupy.

Objection: That desire is not a basic natural trait is in fact independently plausible. It is, endorsed, for example, by certain evaluative theories of desire. According to one such theory, to desire that *P* just is for *P* to *appear to one as good*.[34] Since desire is not purely natural, there is no good reason to think that $\mathbf{Desire}^{\mathbf{U}}$**(NotGood)** is a purely natural proposition. Hence there is no violation of Autonomy.

Reply: Denying the naturalness of desire is, as noted, a problem for the Autonomist, but in addition there is nothing special about desire. We can run a parallel argument for any propositional attitude provided that that propositional attitude can take as objects normative propositions, and is itself the subject of non-trivial normative propositions. Consider the proposition, **Truth**, that all Charles's occurrent beliefs are true. **NormTruth** is the proposition

[33] For ease of exposition I abbreviate the cumbersome "world-time" to "world" throughout the proof.

[34] See (Oddie 2005, 2016 and 2017).

that the normative structure of the world is a **Truth**-norm. **NormTruth** is a purely normative proposition. Consider the proposition that Q is Charles's sole belief: $\mathbf{Believe}^{\mathbf{U}}(Q)$. If belief is natural then $\mathbf{Believe}^{\mathbf{U}}(Q)$ is purely natural. Now we can run a parallel argument to show that that $\mathbf{Believe}^{\mathbf{U}}(\mathbf{NotGood}) \vDash \neg\mathbf{NormTruth}$ is valid. So, if one's strategy to save Autonomy from deep entanglement is to deny that desire is natural, then one will be forced to deny that belief is natural, and indeed to deny naturalness to any propositional attitude that serves as both object and subject of normative truths. It would be something of a disaster if Autonomy could be saved only at the cost of having to deny that just about *any* propositional attitude is natural.

Objection: The entanglement argument may work with any propositional attitude and the attribute of goodness, but perhaps this just shows there is a problem with taking goodness to be one of the basic normative traits. Perhaps what it shows is that the basic normative traits are deontic rather than axiological.

Reply: Let D be any basic deontic feature, or indeed any normative feature, of propositions. For example, $D(Q)$ could be *it ought to be the case that Q*. Let a P-norm be any normative structure that deems a natural structure N to have D just in case N entails P. Let A be any basic natural trait of propositions and let $A^U(Q)$ be the proposition that Q is the unique bearer of that natural trait. $A^U(Q)$ is then a purely natural proposition, for any Q. Let *Most* be the proposition that most of the propositions that have A are true, and let M be the purely normative proposition that is true in a world just in case the normative structure of that world is a *Most*-norm. Let O^D be the proposition that it is not true that the natural structure of the world has deontic feature D. $A^U(O^D)$ is a purely natural proposition, $\neg M$ is a purely normative proposition, and $A^U(O^D) \vDash \neg M$.

Objection: This entanglement argument is really no different from *Dahlia*, which also involves the interaction of propositional attitudes with normative features. So even if this argument is sound it is not new.

Reply: To demonstrate that *Dahlia* is a counterexample to Autonomy, one would have to establish that its second premise **G** is purely natural. $\mathbf{Believes}(P)$ is purely natural, for every P, and **G** is equivalent to an infinite conjunction of propositions of the form $\mathbf{Believes}(P)\supset P$, but this does not ensure that **G** is purely natural. The conditional $\mathbf{Believes}(P)\supset P$ is a hybrid whenever P is purely normative, and conjunctions involving hybrids need not be purely natural. Suppose some cell N_i of the natural partition entails $\mathbf{Believes}^{\mathbf{U}}(\mathbf{O})$, and is compatible with both **O** and $\neg\mathbf{O}$. Let's assume that N_i is also compatible with both **G** and $\neg\mathbf{G}$. $\mathbf{G}^{\mathbf{NAT}}$ assigns true to every every N_i-world. Every N_i-world in which **O** is true **G** is also true, and every N_i-world in which **O** is false, **G** is false. $\mathbf{G}^{\mathbf{CON}}$ assigns true to the former worlds and false to the latter. Thus $\mathbf{G}^{\mathbf{CON}}$ is distinct from $\mathbf{G}^{\mathbf{NAT}}$—i.e. **G** is not purely natural. We could rule out such cells of the natural partition only by assuming that for each N_i that entails $\mathbf{Believes}^{\mathbf{U}}(\mathbf{O})$, either N_i entails $\mathbf{G}\wedge\mathbf{O}$ or N_i entails $\neg\mathbf{G}\wedge\neg\mathbf{O}$. In either case we have to simply *assume* that Autonomy fails (*viz*. either N_i entails **O** or N_i entails $\neg\mathbf{O}$.) So we cannot, without begging the question against Autonomy, establish that the second premise of *Dahlia* is purely natural.

Objection: One must make a distinction between conceptual (or logical) entailment and metaphysical entailment. The Autonomist can happily concede that metaphysical Autonomy fails (and even side with the Naturalist on that score) while insisting that there is no *conceptual* or *logical* entailment between purely natural and purely normative propositions. An overworked example: the identity claim *the morning star is the evening star* is, if true, metaphysically necessary, but it is not conceptually necessary. *X is the morning star* "metaphysically entails" *X is the evening star*, but there is no logical or conceptual entailment involved.[35]

Reply: In my view the highly contested distinction between metaphysical and logical/conceptual necessity is quite unnecessary. The phenomena it purports to explain are better explained by a different theory: namely *role theory*. The claim that *Venus is the morning star* is a contingent claim to the effect that a certain particular (*Venus*) occupies a certain role (*brightest celestial body in the morning sky*). Other particulars (Mercury, Mars, Alpha Centaura, the Space Station) might have occupied that role, and it is a contingent matter, discovered entirely *a posteriori*, that Venus plays the role in fact. Similarly, the claim *the morning star is the evening star* does not identify a particular with *itself* (boring!) but rather claims that two logically distinct roles are contingently co-occupied by one and the same particular. It is by no means logically necessary that they be so. Which particulars occupy which roles is in these, and other typical, cases a contingent affair, and so there is nothing mysterious about the fact that such occupancy claims can only be settled *a posteriori*.[36]

10 Upshot

What I have outlined here is an argument schema for deep entanglement in any framework that features propositional traits amongst both the basic natural traits and the basic

[35] I am grateful to an anonymous referee for pressing this objection. Obviously it is too far ranging a topic to do full justice to in a single paragraph reply, but I sketch my position on it nevertheless.

[36] For an informal introduction to role theory see (Tichý 1987). For the logic of roles see (Tichý 1988).

normative traits. That this kind of argument has until now escaped the notice of metaethicists may be due to the widespread practice of formulating every problem within an artificially restrictive first-order framework, one that simply ignores, *inter alia*, the interaction of propositional attitudes with normative structures.[37] My argument turns on the fact that any framework adequate for representing the relations between the natural and the normative will have to embrace propositional attitudes (like desire, belief, judgement, will, imagine), and those attitudes will not only have to take as possible objects propositions with normative content, but will also be the possible subjects of normative constraints. And that is sufficient for Autonomy to fail.

The non-naturalist normative realist need not be unduly worried by the proof of deep entanglement. It is very far indeed from being a proof of the deepest version of entanglement, that implied by naturalism. And in fact it may come as something of a relief for the non-naturalist realist. It would be rather problematic for the normative realist if the Autonomist were right that absolutely *no* information about the natural state of the world had any logical bearing on the normative structure of the world. For that would beg difficult questions about how creatures such as ourselves could ever get an epistemic purchase on the normative truth. [38]

References

Belnap ND, Steel TB (1976) The logic of questions and answers. Princeton University Press, Princeton

Brown C (2014) Minding the is-ought gap. J Philos Log 43(1):53–69

Jackson F (2003) Cognitivism, a priori deduction, and Moore. Ethics 113:557–575

Lewis D (1988) Statements partly about observation. Philos Pap 17(1):1–31

Mackie JL (1946) A refutation of morals. Australas J Philos 24:77–90

Mackie JL (1977) Ethics: inventing right and wrong. Penguin Books

Mares ED (2010) Supervenience and the autonomy of ethics: yet another way in which Relevant Logic is superior to Classical Logic. In Pigden 2010:272–289

Nelson MT (1995) Is it always fallacious to derive values from facts?. Argum 9(4):553–562

Oddie G (1986) Likeness to truth. Reidel, Dordrecht

Oddie G (2005) Value, reality and desire. Oxford University Press, Oxford

Oddie G (2016) Fitting attitudes, finkish goods, and value appearances. In: Schafer-Landau Russ (ed) Oxford Studies in Metaethics. Oxford University Press, Oxford, pp 74–101

Oddie G, Demetriou D (2007) The fictionalist's attitude problem. Ethical Theor Moral Pract 10(5):485–498

Pigden C (1989) Logic and the autonomy of ethics. Australas J Philos 67(2):127–151

Pigden C (ed) (2010) Hume on is and ought. Palgrave Macmillan, New York

Prior AN (1960) The autonomy of ethics. Aus J Philos 38(3):199–206

Restall G, Russell G (2010) Barriers to implication. In Pigden 201:243–259

Schafer-Landau (**forthcoming**). Desire and the good: finding the right fit. In Deonna J, Lauria F The Nature of Desire. Oxford University Press (**forthcoming**)

Schurz G (1997) The is-ought problem: a study in philosophical logic. Kluwer, Dordrecht

Tichý P (1986) Constructions. Philos Sci 53(4):514–534

Tichý P (1987) Individuals and their roles. In: Svoboda V et al (eds) Pavel Tichý's collected papers in logic and philosophy. Filosofia and the University of Otago Press 2004, Prague and Dunedin, pp 749–763

Tichý P (1988) The foundations of Frege's logic. De Gruyter, Berlin

[37] This is how (Brown 2014) squeezes Autonomy out of an analysis similar to mine.

[38] I would like to thank the audience members at various presentations of this paper for comments: those at the Prior conference in Christchurch in December 2014; those who attended my talk at the Uppsala Philosophical Society in October 2015; department members at Notre Dame University colloquium in November 2016. I also thank Dom Bailey and two referees at *Topoi* for many helpful criticisms and corrections.