



2.4 | When Falsification Strikes: A Reply to Fodor

Leda Cosmides and John Tooby

We thank Jerry Fodor for taking the time to write his reply. In it, fortunately, he has become clear enough that we can identify the simple factual errors on which he bases his arguments, as well as several other sources of his theoretical and experimental confusions. Once these are resolved, perhaps Fodor will be able to identify actual weaknesses in our experimental program that we could profit by addressing. Indeed, over the last two decades, we have made it a policy to test systematically against every coherent counterhypothesis to social contract theory and hazard management theory. At present, as far as we are aware, there are no remaining viable alternative theories that have not been clearly falsified by the accumulated body of results (see Cosmides & Tooby, 2005a, and this volume, for review). Obviously, testing different hypotheses requires experiments with different controls. For this reason, it is a feeble form of criticism on Fodor's part to claim that a single experiment designed to test hypothesis A, when taken in isolation, does not test against hypothesis B, when there already exist other widely known experiments that have falsified hypothesis B.

The important but uninteresting point is that every empirical test that Fodor (this volume; 2000) claims we should have made in order to establish our case has already been conducted. These tests are reported in papers that Fodor cites, as well as in papers that he does not cite. In all but a single case, he offers no argument to dispute their outcomes. He simply ignores their existence, arguing for a counterhypothesis that has been repeatedly falsified.

A more important—and more interesting—point concerns the unexamined assumptions inherent in Fodor's simple and traditional opposition of “logical form” with “content.” We will turn to the problem with that opposition in due course.

Fodor's First Fundamental Error

In his reply Fodor correctly states that “C&T sought to test the hypothesis that inference is content specific in a perfectly sensible way: namely, by showing that, depending on the content of the materials, formally identical inferences exhibit different interactions with performance in the Wason selection task” (p. 137). He then, unfortunately, makes a fundamental factual error and builds his countertheory around it. Fodor goes on (emphasis added): “*The experimental manipulation in the C&T experiment contrasted deontic conditionals with descriptive ones . . .*” (pp. 137–138).

Even in the original article reporting social contract experiments (Cosmides, 1989), there were *nine* experiments, rather than one. These experiments contrasted social contract conditionals with a variety of deontic conditionals that were not social contracts, and not (as Fodor mistakenly believes) solely with descriptives. Fodor elaborates his fanciful view that we only contrasted “deontic conditionals with descriptive ones” (pp. 137–138), creating “an unresolved confound of content differences with logical form differences” (p. 139). He continues, “(N)otice that observed interactions of content variables with performance are germane to the domain specificity of inferential processes only if the logical form of the inferences is held constant in the materials manipulations. I argued that, as a matter of fact, this was *not* the case with respect to C&T’s test materials since (according to me), descriptive and deontic conditionals ipso facto differ in their logical form” (p. 138). Fodor builds his entire case on an imaginary confound in which social contract rules (which are deontic) are only contrasted with descriptive rules (which are nondeontic). For example, Fodor (2000, p. 29) states that our interpretation of our experimental results “depends critically on assuming that deontic conditionals and their indicative controls are identical in structure.”

Yet, for two decades we (and others) have systematically tested not only descriptive (indicative) control rules but control rules drawn from a large variety of theoretical categories, including a large number of conditional rules that have exactly the logical form that Fodor claims is actually responsible for the performance effect we misattribute to social contract content (i.e., deontic rules). If Fodor’s substantive claim were true (i.e., that high performance derives from the logical form deontic rules have), then subjects should perform well on deontic rules as a class, and the presence of social contract content in the deontic rule should be irrelevant. Unfortunately for Fodor’s theory, it has been shown on numerous occasions that being a deontic rule is not, by itself, sufficient to elicit high

performance. That is, the robust and striking performance differences that subjects exhibited between deontic conditionals that are social contracts and deontic control conditionals that are not social contracts cannot be explained as the result of the non-social-contract deontic rules being descriptive rules, because (as Fodor emphasizes) deontic rules are not descriptive rules (see below, and Cosmides & Tooby, this volume, for a review of these experiments). Fodor's error could not be more basic, and it is hard to credit that he could actually have read any of the experimental findings that he cites and is attempting to construct theories about.

Fodor's Second Fundamental Error: Logical Form and Content Are Mutually Exclusive Hypotheses

Fodor is admirably clear in stating another of his major premises, which again allows us to pinpoint his confusion. He states (this volume): "As far as I can tell, C&T have lost their grasp of the dialectical situation. What they need—the evidence that would be relevant to whether inferential processes are domain specific—is this: *there is no effect of logical form on inferential processes when content is controlled*. But the evidence they've got is only this: *there are effects of content on inferential processes when logical form is controlled*" (p. 140).

This is, of course, wrong on its face. Accepting, for the moment, Fodor's simple opposition between the effects of logical form and content, there are (at least) four possibilities: logical form could have an effect, but not content; content could have an effect, but not logical form; neither might have an effect; or both could have effects. *Nothing about the claim that there exist domain-specific inferential procedures requires that there be no effects of "logical form."* These are logically independent claims. To support social contract theory, it is sufficient to show that (1) there are principled domain-specific effects, and (2) these cannot be explained as the (supposed) effects of logical form, whatever they might be. We have in fact (1) predicted and experimentally confirmed the reality of a large number of independent, previously unknown, highly specific content effects and (2) shown these effects cannot be explained as the effects of logical form (because, for one thing, logical form was held constant while the effects appear and disappear with content manipulations—exactly the tests Fodor said were "germane"). Importantly, these novel effects were predicted based on a prior principled functional analysis of what social exchange inference procedures would have to be like in order to carry out their evolved function. Indeed, we have been able to predict and to produce in a principled

way distinct patterns of card choices that were “illogical” (according to the propositional calculus) that no one had ever observed before. It is simply a straightforward logical error to claim that we must also show that there are no effects of logical form.

It is possible that Fodor derives this belief from his biologically problematic views on mandatory information encapsulation in evolved specializations. If so, Fodor may think that the claim that social contract procedures are operating on representations entails the claim that other, more general logical operators cannot also operate on them. That is, if representations are “in” a social contract module, they cannot be affected by other aspects of the architecture. But nothing about an evolutionary functional perspective on neurocomputational design leads to the conclusion that all evolved computational procedures will be completely information encapsulated. In an evolutionarily well-engineered brain, circuits should be networked together whenever they (over evolutionary time) can pass useful products to each other, and such interactions do not impede efficiency. That is, information encapsulation should be only an occasional outcome and only occur when the benefits of isolation outweigh the costs. It seems unlikely that this is the case in reasoning, in which information sets are potentially relevant to each other (Cosmides & Tooby, 2000b). This suggests that various inferential operators (specialized and more general) will act on the same processing stream. No one has made any argument about why natural selection, acting in a content-structured world, would act to preserve the abstract purity of a set of general logical procedures by blocking their supplementation with additional content-sensitive problem-solving specializations. They should all be interacting in fine-grained ways in reasoning processes—a proposal that accounts for the rich, content-sensitive character of human reasoning, as well as the routine, if occasional, exhibition by mortals of inferential chains that seem to reflect the expression of more general logical operators.

Fodor Ignores the Empirical Implications of the Hypothesis that, in a Multimodular Mind, Content Is an Element of Logical Form

More interesting issues are raised by the unexamined assumptions inherent in Fodor’s opposition of “logical form” with “content”. He accepts it as obvious that content and logical form are different ontological kinds. However, this view is only true for domain-general logics, in which content is ignored by operators designed to act on any proposition p , regardless of content. The proposal for content-sensitive reasoning systems is far more

radical than this, because it breaks down the distinction between logical form and (some) content. For domain-specialized inferential systems, *content is an element of logical form*.

In our view, the human neurocomputational architecture contains a number of functionally distinct inferential systems. Evidence suggests that some are relatively domain general and include procedures that embody certain elements of the predicate calculus, such as negation, variable binding, and *modus ponens* (Rips, 1994; Cosmides & Tooby, 1996b and this volume, sections I.ii, II.i.i.i; all further citations by section refer to Cosmides & Tooby, this volume). Yet, surprisingly, the architecture seems to lack the ability to reliably deploy many other basic logical operations. For example, the human mind seems to lack *modus tollens*, which makes following arguments involving Popperian falsification a nonroutine mental achievement for scientists as well as civilians. Thus, *contra* Fodor, we have always accepted the existence of (some) general logical procedures and consider it possible for there to be some effects of what Fodor calls (rather vaguely) “logical form” (although experiments show them to be—at most—minor in magnitude, compared to social contract effects).

Alongside these domain-general reasoning procedures are (we argue) a number of domain-specialized inferential systems. Among these is a functionally distinct system specialized for interpreting and reasoning about social exchange interactions—the *social contract algorithms*—and a system specialized for reasoning about precautions and hazards. According to social contract theory:

A. This system is equipped with domain-specialized rules of transformation and inference that operate over abstract yet contentful primitives (*agent x*, *benefit to agent x*, *cost to x*, *requirement of y*, *entitlement*, *obligation*, etc.). That is, it has a domain-specialized syntax, which implements and applies a domain-specialized logic of social exchange. The rules of transformation and inference implemented by the social contract algorithms differ from those of the predicate calculus in two ways.

First, they operate over abstract representations of *contentful* conceptual primitives (*benefit to x* being a particularly important one). Consider, for example, this conditional: “If a man is eating cassava root, then he must have a tattoo on his face.” In predicate logic, it is irrelevant whether “eating cassava root” is a benefit or a cost to the eater, because its syntax represents this activity merely as proposition *P*. But when the mind interprets “eating cassava root” as a benefit to the eater, that representation—*benefit to agent x*—is a *syntactic* element in the logic of social exchange, an element of

logical form with respect to the syntax of its domain-specialized internal logic. If cassava were a poison, this activity would be represented as a *cost to agent x*, so the rule would no longer have the same logical form *with respect to* the logic of social exchange. In short, what counts as content and what counts as logical form will be determined differently by the predicate calculus, by the logic of social exchange, by the logic of hazard management, and so on, according to the procedures embodied in each system.

Second, and crucially, the rules of transformation in the social exchange system license inferences that are appropriate for social exchange *but that are invalid under the predicate calculus*. Because they are specialized for exchange, they do not capture the inferences we make about indicative conditionals. Nor do they capture the inferences we make about deontic conditionals involving precautions, the inferences we make about threats, or the inferences we make about deontic conditionals that are neither social exchanges or precautions. Because they make correct inferences only within the domain of social exchange, social contract algorithms are not designed to be activated or applied outside of their proper domain.

B. Because social exchange inferences are invalid outside the domain of social exchange, they are designed to be activated only to the extent there are cues suggesting the potential presence of a social exchange situation. (Parametric investigation suggests that cues activating social contract algorithms are additive (section IV.vii.i), so that the more cues, the stronger the social contract effect.) Social contract cues are elements that indicate the presence of a situation fitting the template for a social contract: "If you take benefit B, then you must satisfy the requirement R of the agent that controls benefit B" (section II.i.). It follows directly from this that relevant cues include (1) logical operators in the conditional rule that, in context, clarify that it is a rule in which an *agent* is *requiring* something of another *agent* (e.g., *must*, *It is required that*, etc.), and (2) terms that, in context, are interpreted as benefits, costs, agents (etc.) congruent with the template.

Fodor puzzlingly claims that, for social contract theory to be correct, we must show that logical form has no effect. In reality, however, social contract theory straightforwardly predicts (based on A and B above) that social exchange reasoning will be sensitive not only to the presence or absence of social contract content but also to manipulations of social contract theory's version of logical form. One element of this happens to coincide with the element of general logical form that Fodor thinks is responsible for evoking striking performance effects.

Fodor thinks high levels of violation detection will be elicited by any deontic conditional because their logical form implies that these rules are

really about “requiring Q,” and anyone can see that cases of *not-Q* violate this requirement. That is, what Fodor has extracted from the social contract account to advance as the key element of his deontic account is a *requirement* (implicitly or explicitly made by an agent) on other agents. This is one-half of the social contract template and so is one potential cue to a social contract. All Fodor is lacking is the other half of the social contract template, which specifies that meeting the requirement allows access to the rationed benefit. *So both theories predict that adding cues about an agent making a requirement should have an effect.* Because Fodor labels this “logical form,” *both theories predict that logical form* (in this delimited sense) *should have an effect* (though for different reasons). Fodor claims that this element is responsible for the entire (strong) effect on *not-Q* selections that is misattributed to social contract theory. In contrast, social contract theory predicts that this manipulation, by itself, should produce a weak effect on *not-Q* selections (assuming nothing else blocks a social contract interpretation), but when combined with the missing *rationed benefit* component of the template, the enhanced violation detection effect should be strong. Which claims are correct?

Experimental results closely track the predictions of social contract theory (see table 2.4.1). For example, experiments explicitly designed by Beaman (2002) to test Fodor’s claims show that there are indeed effects of using requirement language—as both theories predict. However, the theories diverge sharply on their predictions about the size of the effect (among other things). Unlike Fodor’s proposal, social contract theory predicts the effect will be small, because the agent-requirement cue is only one half of the social contract template. In order to provide a counterexplanation to social contract theory, Fodor’s account requires the effect to be as large as the performance boosts otherwise attributed to social contract manipulations. In reality, the effect of requirement language is a mere 15 percentage point jump (hop?) in *not-Q* selections—something Fodor nonquantitatively treats as a vindication of his theory. Fodor not only ignores the fact that the magnitude of the effect of logical form is far too small to pose a challenge to social contract theory, but he completely ignores the far more striking effect of Beaman’s adding back in the missing half of the social contract template: the rationed benefit component. That is, according to social contract theory, requirement language should not boost performance to the high levels typical of social contracts unless another key syntactic element is present: something that can be interpreted as a *benefit* to the agent of whom *Q* is required.

Table 2.4.1
Tests of Logical Form Versus Social Contract Content

	Predictions		Results	
	% selecting not-Q card		% choosing:	
Logical Form (LF) of rules = indicative or deontic?	Fodor	SCT	not-Q card	P & not-Q
Beaman (2002)				
1a. If a card has a vowel on one side, then it has an even number on the other side.	Low	Low	25 (odd #)	Not reported
2a. If a person is under 18, s(he) drinks coke in this pub.	Low	Med-high	65 (beer)	Not reported
	LF = indicative			
1b. It is a requirement of yours that if a card has a vowel on one side, then it has an even number on the other side.	High	Low-med	40 (odd #)	Not reported
	(SCT: LF = ambiguous: indicative or deontic)			
2b. It is a legal requirement that if a person is under 18, s(he) drinks coke in this pub.	High	High	90 (beer)	Not reported
	LF = deontic			
	1b = 2b	1b ≪ 2b		
	LF = deontic			
Cosmides & Tooby	LF = deontic for all			
3. It is required that if someone is under 18, that person drinks coke.	Fodor	SCT	not-Q	P & not-Q
not-Q = drinking beer (Q = drinking coke, P = 16 years old, not-P = 30 years old; N = 33)	High	High	75.8 (beer)	63.6
4. It is required that if someone is under 18, that person drinks coke.	= Rule 3	< Rule 3	< Rule 3	43.8
not-Q = drinking milk (Q = drinking coke, P = 16 years old, not-P = 30 years old; N = 32)			53.1 (milk)	
5. If a person is drinking beer, then that person must be over 18.	≤ Rule 3	> Rule 3	> Rule 3	85.2
not-Q = 16yrs old (Q = 30 years old, P = drinking beer, not-P = drinking coke; N = 27)			96.3 (16yrs)	

The results shown in table 2.4.1 are from experiments using the Wason selection task; points where Fodor's predictions contrast with social contract theory (SCT) are in boldface. Beaman (2002) conducted a parametric test examining how adding "It is required that" before a conditional would affect not-Q selections for two rules: one with social contract content and one without. Fodor (2000) views rules [1a] and [2a] as indicative conditionals that should elicit low levels of not-Q selections, but rules [1b] and [2b] as deontic conditionals that should elicit high levels of not-Q selections. Social contract theory (SCT) also views the vowel rule [1a] as indicative, but sees rule [2a] as ambiguous as to interpretation. Beaman's subjects live in a society where alcohol consumption is a privilege reserved for adults, i.e., a rationed *benefit*. Rule [2a] is set in the context of a pub serving only whisky or coke, so most college students probably interpret it as a (weird) paraphrase of the well-known drinking age law (rule [5]), which is a (deontic) social contract. But [2a] could also be seen as indicative, describing the tastes or habits of people under 18.

Requirement language rules out indicative interpretations for both rules. SCT predicts an improvement in performance for [2b]: Anyone who was tempted to interpret [2a] as merely describing teen-agers' habits would now understand that the rule is meant to express a deontic social contract, the drinking age law. "It is a legal requirement that" *means* there are agents requiring something of other agents (part of the syntax of social exchange), and this requirement must be satisfied to be eligible for a *benefit* (another key syntactic element of social exchange). In contrast, although rule [1b] has requirement language and the instruction to look for cards that "breach your criteria" (implying someone else was supposed to follow your requirement—hence cues that an *agent* is *requiring* something of another *agent*), it has nothing interpretable as a *benefit*. Thus, SCT predicts far higher performance on the deontic social contract [2b] than on the deontic vowel problem [1b], which was found. Indeed, *not-Q* selections on the deontic vowel problem were no higher than what is usually found for *indicative* conditionals (typically, ~50% for indicatives, with ~25% choosing P, not-Q, and no other cards; see figure 2.4.1). The problems we did—rules 3–5—all involve deontic conditionals. Fodor's theory predicts they will all elicit equally high levels of *not-Q* selections (or perhaps fewer for rule [5], see text). In contrast, SCT predicts [5] > [3] > [4], which is the actual observed pattern.

As table 2.4.1 shows, that is precisely what Beaman (2002) found: 90% *not-Q* selections for the deontic social contract compared to 40% for the deontic vowel problem (which follows only half of the social contract theory template). These data sharply violate Fodor's key prediction: that requirement language is sufficient to elicit high levels of *not-Q* selections. As Beaman shows, holding logical form constant, the addition of the rationed benefit content caused an increase in performance, of 40 and 50 percentage points (rules [1a] vs. [2a] and [1b] vs. [2b], respectively), dwarfing the effect of "logical form." This social contract content effect is predicted by social contract theory but not by Fodor's deontic/indicative account, and it corresponds to an effect size (ϕ) of .73, more than double the effect size of .30 found for requirement language. That is, Beaman has replicated the results of our benefits experiments (section IV.vi and below) using Fodor's own preferred manipulations. Even worse for Fodor's account, the social contract *without* requirement language, [2a]—which Fodor (2000) views as indicative—outperformed by 25 percentage points the *deontic* (but non-social-contract) vowel problem, [1b], that on Fodor's account should elicit high performance. (The social contract elicited 65% vs. 40% for the non-social-contract deontic rule.) Beaman himself concludes that Fodor's variant of the Wason selection task "is insufficient to account for the 'cheater detection' effect" (in the abstract, cunningly concealed from Fodor).¹ Yet from reading Fodor's reply, readers equipped with normal pragmatics would conclude that the Beaman experiment supports Fodor's proposal that logical form manipulations (with a 15 percentage point effect) explain away social contract effects (which add an additional 50 percentage points).²

To prevent Fodor from improvising another post hoc explanation for Beaman's results, we conducted an additional experiment designed to test Fodor's proposal. This experiment compares two *identical* deontic rules, both of which were Fodor's (2000) flagship example of a *not-Q*-eliciting deontic conditional: "It is required that if someone is under 18, then that person drinks coke" (rules [3] and [4] in table 2.4.1). The rules are identical; all that changes is whether the *not-Q* card refers to beer or milk. This should make no difference on Fodor's theory but constitutes the addition or removal of the rationed benefit component of the social contract template and so should cause a substantial change in subject performance. When *not-Q* refers only to beer (or any kind of alcohol), then this rule is logically equivalent to "It is required that if someone is under 18, that person does not drink alcohol." Like Beaman's, our subjects see drinking alcohol as an age-related privilege, so they should interpret this rule [3] as

restricting access to a *benefit*. But this interpretation should be blocked when *not-Q* refers to milk. Rule [4] is logically equivalent to “It is required that if someone is under 18, that person does not drink milk”: this is certainly deontic, but it does not fit the input conditions for a social contract very well. Drinking milk is not seen as a privilege relative to drinking coke (most children think the opposite), nor should any mammal see milk drinking as a benefit reserved only for adults. We gave subjects these two problems, and the results (table 2.4.1) are clear: when *not-Q* = beer, supporting a social contract interpretation, performance was 20 percentage points higher than when *not-Q* = milk (indeed, the milk version elicited no more *not-Q* selections (53%) than indicative conditionals do; see figure 2.4.1).

And how does Fodor’s flagship example compare to a version of the rule whose logical form more closely fits the syntax of social exchange? Rule [5], the drinking age law as standardly given in the literature,³ is the contrapositive of rule [3], Fodor’s flagship deontic rule. But rule [5] better fits the benefit requirement template of a social contract: the benefit—drinking beer—is mentioned *in* the deontic rule, which says what is required if one is to be entitled to this benefit. It should outperform Fodor’s rule [3] because rule [3] is more difficult to recognize as a social contract: to derive a rule fitting the benefit-requirement template, one must interpret rule [3] as implying rule [5], its contrapositive. As social contract theory predicts, the standard drinking age problem, rule [5], outperforms Fodor’s flagship rule [3] by about 20 percentage points—see table 2.4.1. But this is the *opposite* of what Fodor’s argument predicts.

Even worse, Fodor rejects the view that rule [3] implies its contrapositive, rule [5], because he thinks you cannot be required to be over 18.⁴ Given rule [5], subjects should be puzzled; to succeed, they would have to infer that the experimenter was confused and really meant to state Fodor’s rule [3]. This implies that Fodor’s rule [3] should outperform our rule [5] (which it does not). Or, if subjects are surprisingly adept at deriving this *modus tollens* inference from rule [3], the two rules could elicit equivalent performance (which they do not). But there is nothing in Fodor’s account that can explain what we found: *Better* performance on rule [5] than on Fodor’s optimal logical form, as expressed in rule [3].

These examples provide an opportunity to understand how far we have come from Fodor’s formulation of the content-logical form distinction. If the cognitive architecture contains many different inferential systems, then one cannot make the same distinction between logical form and content effects, and his distinction dissolves. Putting “it is required that”

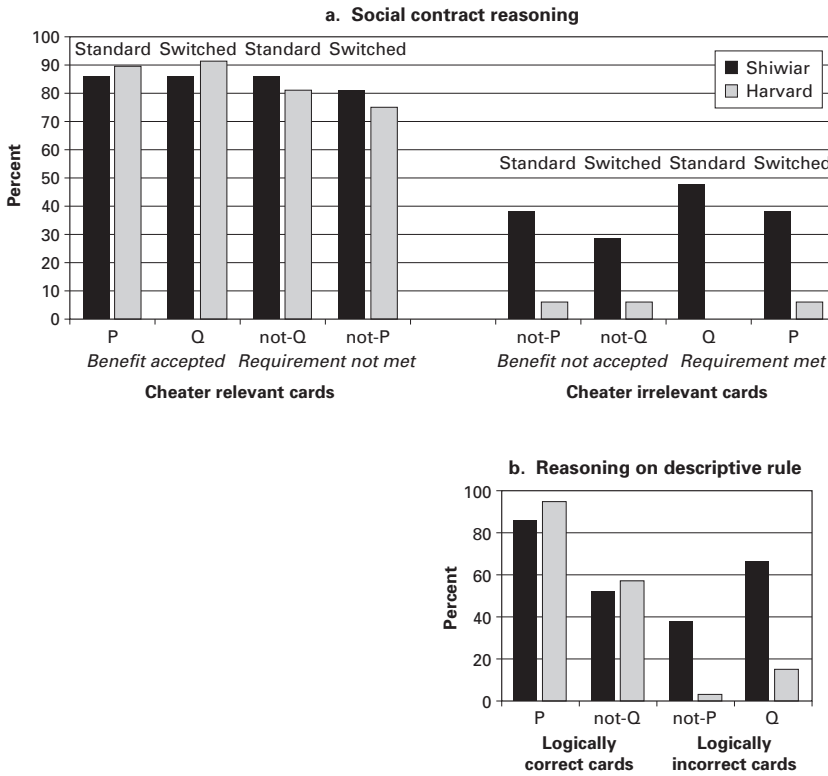


Figure 2.4.1

Panel a. The logical form of a social contract can be changed from standard to switched, yet cheater relevant cards are chosen equally often (whether subjects are Harvard students or Shiwiar—indigenous residents of the Amazon). This is a case of holding content constant, changing logical form, and showing that this does not affect the frequency with which “requirement not met” cards are selected. Panel b shows that *not-Q* is selected about 50% of the time for familiar indicative rules. This is important for evaluating whether deontic rules elicit more *not-Q* selections than indicatives; see results for Fodor’s flagship rule when the *not-Q* card refers to milk rather than beer, table 2.4.1. (From Sugiyama, Tooby, & Cosmides, 2002.)

in front of a rule that would ordinarily be interpreted as an indicative conditional is not just a change in logical form; it is also a change in *content* relative to the syntax of social exchange. It implies that there is a situation with an *agent* requiring something of another *agent*—a partial but incomplete cue of social exchange. Similarly, experiments that add or subtract a rationed benefit linked to the requirement manipulates a *syntactic* element—syntactic from the point of view of the representational primitives of the social exchange system (although not syntactic within predicate or deontic logic).

In short, in evaluating the evidence, it is important to keep in mind the similarities and differences between logical form as defined by the syntax of social exchange and logical form as argued by Fodor (using his personally improvised synthesis of the predicate calculus as applied to deontic rules).⁵ Because social exchange syntax is defined over contentful representations (such as *benefit to x*), it follows that reasoning will be sensitive to aspects of a problem's content that matter to this syntax. Content manipulations targeted to this syntax should change how the mind represents a given sentence, transforming its interpretation from a deontic conditional that is not a social contract to one that is (section IV.vi), and from the logical form of an indicative conditional to the representation of a deontic social contract (sections IV.ii and IV.iii.i).⁶

Précis of Other Evidence Falsifying Fodor's Proposal

When social contract theory was first formulated over two decades ago, we proposed that indicative conditionals and (deontic) social contracts are interpreted differently and processed via different "inferential routes" (Cosmides, 1985; Cosmides & Tooby, 1989). We were therefore surprised when Fodor (2000, this volume) proposed the same thing, in the belief that this claim contradicts and refutes social contract theory. However, our initial proposal went far beyond proposing different inferential routes for indicatives and deontics: we proposed that (deontic) social contracts are processed via a different inferential route than other *deontic* conditionals.

Fodor's (2000) proposal contrasts sharply with this latter claim. Empirically, his argument comes down to this: *having the logical form of a deontic conditional is sufficient to elicit the high levels of violation detection found for social contracts; this (purportedly strong) effect of being a deontic rule explains away the evidence previously and erroneously taken as support for social contract theory.*

Fodor's proposal that high levels of violation detection are produced by any deontic conditional was reasonable two decades ago, when we began our research. That is why our studies have, from the outset, included conditions that compare deontic conditionals that are social contracts to other deontic conditionals (for review, see sections IV.v–IV.viii). Fodor's proposal is far less reasonable now, when such a theory would have to address the dozens of findings that systematically falsify its key predictions. Fodor's claimed "confound" between content and logical form is absent in the tests in which deontic logical form is held constant and content is varied in ways relevant to the proposed syntax of social exchange. If Fodor's proposal were correct, all of these deontic conditionals should have elicited high levels of violation detection. But they do not. *Deontic conditionals that are not social contracts (and not hazard/precaution rules) do not elicit high levels of violation detection on the Wason task.*

Our many experiments along these lines were conducted not to flog a long dead horse—that violation detection is the result of deontic logical form. That theory is long dead, even if Fodor is now struggling to saddle up the corpse and ride triumphantly off on it. Instead, by holding deontic logical form constant and varying elements relevant to the syntax of social exchange, we were able to test (and confirm) key predictions of social contract theory.

None of these manipulations should implode performance if Fodor were right, because his proposal empirically predicts that it is the property of being a deontic conditional that elicits high levels of violation detection: (1) For deontic rules, it should not matter if *P* is a *benefit to agent x*, a *cost to x*, a *hazard to x*, and so forth (yet it makes a dramatic difference; sections IV.vi and IV.viii) (2) it should not matter whether the violation was intentional or accidental (yet it does; see section IV.vii). And (3) the violations people detect should be of the requirement specified by the deontic conditional they were given. They should not be looking for violations of a requirement specified by a *different* deontic rule, a rule that can be derived from the conditional given only by applying domain-specialized inference rules that are invalid outside the domain of social exchange (yet they do; see section IV.iii).

Experiments holding deontic logical form constant include the following, discussed in our paper (this volume):

- Section IV.vi: *Benefit experiments* These varied whether *P* is a *benefit to x* or a *cost to x*. Violation detection dropped precipitously when the deontic rule no longer regulated access to a *benefit*—by 36 percentage points in the

example discussed. This drop was predicted in advance by social contract theory: deontic rules that do not regulate access to benefits do not fit the syntax of social exchange. As discussed, Beaman (2002) found similar results: holding deontic logical form constant, 90% of subjects chose *not-Q* for the social contract versus only 40% for the non-social-contract rule.

- Section IV.vii: *Innocent mistake experiments* These held constant both logical form *and* social contract content, so, on Fodor's account, they would necessarily be assigned the same logical form and interpretation. All that was varied was whether potential violations were intentional or accidental. The function of the cheater detection system is to identify cheaters whose errors benefit themselves (and who are disposed to cheat again)—not those who commit innocent errors. As predicted, violation detection was high for intentional cheating, but not for innocent mistakes (yet these are identical rule violations, differing only in whether they could have been intentional). In contrast, the function of the precautionary system is to identify people in danger, which can happen accidentally or intentionally. As predicted, the same accident-intention manipulation did not affect performance on deontic precautionary rules at all.

- Section IV.viii: *Neural dissociation between deontic conditionals* There are two classes of high-performing deontic conditionals that have been identified so far: social contracts and precautionary rules. On Fodor's account, both would be processed via the same "inferential route." Yet brain damage can leave the ability to detect violations of deontic conditionals that are precautionary intact, while selectively impairing the ability to detect violations of deontic social contracts (Stone et al., 2002). These two classes of rule differ only in whether they fit the syntax of social exchange or the syntax of the hazard/precaution system—but Fodor's account denies that these content-specialized syntaxes exist. If all that exists is domain-general logical machinery, then when it is impaired, performance on all rule types should be degraded. It is not. Moreover, functional magnetic resonance imaging research confirms that what we consider two kinds of conditionals, and Fodor considers a single kind of conditional, are in fact treated as two distinct classes by the mind (Ermer et al., 2006). Fodor's response to this increasing body of cognitive neuroscience findings showing the mind distinguishes (along proposed evolutionary lines) classes of rules that Fodor lumps into a single class reads, in full: "Who except a dualist would deny it?" (p. 140).

- Section IV.iii.ii: *Perspective change* In these experiments, precisely the same deontic social contract—same logical form, same content—was used in both conditions; all that varied was which agent was the potential

violator. Performance changed in precisely the way social contract theory predicts. For this to happen, subjects had to infer an implication of the deontic conditional that is valid for social exchange but invalid under any domain-general logic (deontic or predicate logic).

▪ Section IV.iii.i: *Switched social contracts* These experiments held constant (1) social contract content, (2) the potential violator, and (3) logical form *as defined by the syntax of social exchange*. All that varied was logical form *as defined by domain-general logics* (deontic or predicate logic). This change in domain-general logical form did not influence cheater detection: subjects were just as good at looking for those who did not satisfy the requirement even when this corresponded to *not-P* (rather than *not-Q*). It did, however, cause subjects to strikingly violate the valid operations of domain-general logics, falsifying Fodor's idea that it is a general logic that is doing the work.

Fodor's Response to Experimental Falsification

What does Fodor say in his reply when confronted with the large body of empirical findings that falsify his view (identified in the paper Fodor is replying to)? Curiously, he does not respond to the great majority of them. Instead:

1. Fodor first reiterates the error in his 2000 paper, stating "The experimental manipulation in the (sic) C&T experiment contrasted deontic conditionals with descriptive ones" (pp. 137–138). It was based on this error that he claimed all of our experiments confound logical form and content and, therefore, provide no evidence for a cheater detection mechanism.
2. Fodor then contradicted this statement by acknowledging one experiment (of the many we discuss) in which deontic logical form was held constant and only social contract content was varied: a single experiment manipulating *benefit*. In his reply, Fodor began by stating that "observed interactions of content variables with performance are germane to the domain specificity of inferential processes only if the logical form of the inferences is held constant in the materials manipulations" (p. 138). So this experiment manipulating the degree of benefit should be germane: it held logical form constant ("requiring Q" of anyone who does P), varying only whether P was a *benefit to x* (going out at night) or a *cost to x* (taking out the garbage at night). That the former (a social contract) elicited good violation detection from 80% of subjects, compared to only 44% for the other deontic conditional, would seem to be a strong violation of Fodor's

claim that deontic logical form is the factor that elicits high levels of violation detection.

Fodor makes three contradictory responses: (1) he traces channels his grandmother, saying that she could have predicted this content effect; (2) he says that it is irrelevant to hold logical form constant and vary content (after previously admitting that this is indeed “germane”), because all that shows is a content effect, and (3) he says (erroneously) that we predict reasoning will be insensitive to logical form. Without saying why—and in direct contradiction to social contract theory as laid out in the paper to which he is replying—he states that social contract theory predicts that changes in logical form will have no effect on reasoning when (explicit) social contract content is held constant. We have already dealt with (3), but we will take the other two in turn.

Fodor’s Grandmother

Those familiar with Fodor’s writing know that he usually resurrects his grandmother when he wants his intuition to do the work that a good computational theory should. In this case, Fodor’s improvised account is motivational: subjects “are more interested in understanding the requirements for doing things that they might want to do (going out for the evening) than they are in understanding the requirements for doing things that they don’t want to do (putting out the garbage)” (p. 141). Here as elsewhere, Fodor makes no attempt to consider whether empirical predictions entailed by his ad hoc arguments are contradicted or supported by any finding outside of the inconvenient result he wishes to explain away. Unfortunately for the reputation of Fodor’s grandmother, this explanation is false. That is, a large number of social-contract and non-social-contract rules have been tested in the past twenty years, and performance does not vary as a function of what *subjects* want: As social contract theory predicts, performance is just as high for dull, boring social contracts that specify requirements for doing things our subjects have no interest in, from acquiring potatoes or ostrich eggshells to procuring corn (see, e.g., Fiddick et al., 2000; Cosmides, 1989). What does govern performance is whether the rule restricts access to something that *the character in the story wants*—to something that the character considers to be a *benefit*. It does not matter whether this is something that the *subject* might want to do or finds interesting.

Fodor’s motivational explanation is equally undercut by the innocent mistake experiments. Here, the social contract is precisely the same in both conditions, so *there can be no differences* in how much subjects might want

to do *P* (in this case, to go to a good high school). Yet performance dropped for accidental violations, but not for intentional violations. The motivational account cannot be rescued by positing that subjects are more interested in intentional than accidental violations of deontic rules because this same manipulation has *no effect* on deontic rules that are precautionary. Motivation cannot even explain an effect of intentionality that is specific to social contracts. Fodor's grandmother would surely know that situations in which people do not get what they are entitled to are interesting, but accidental violations create this interesting situation as surely as intentional ones do. Yet performance is high only when this violation is produced by a cheater (section IV.vii; Cosmides & Tooby, 2005a, pp. 609, 618, 621).

It's "Just" a Content Effect

Fodor's next line of defense is to dismiss the idea that content effects *per se* can be informative about these issues, because introductory textbooks routinely admit that content affects cognition. This might be a challenge if there were an alternative theory that economically predicts the same large and diverse set of content effects on reasoning that we have found. But no such theory exists. Nevertheless, this approach can be shown to fail even in principle.

Assume, as Fodor does, that the cognitive architecture contains only one domain-general inference system, something like the predicate calculus. If this were true, it would be possible to sharply distinguish between effects of logical form and effects of content: some content might boost the application of the predicate calculus, much like (to use Fodor's example) adding teddy bears rather than integers boosts the application of the cognitive number system for preschoolers. But social contract theory is not about cheater content boosting the application of predicate logic (or any domain-general deontic logic). It proposes something far more radical: that a *logic of social exchange* inhabits the cognitive architecture, alongside other content-dependent, domain-specialized syntactic systems, and operates according to its own procedures, which sometimes sharply diverge from inferences licensed by predicate logic. This is exactly why the perspective change and switched social contract experiments are so informative: the pattern of results they elicit depends on subjects' spontaneously making inferences that are licensed by social contract algorithms, but that are *invalid* in predicate logic, and *invalid* in a general deontic logic. Any explanation about facilitating the use of predicate logic by fleshing it out in some kind of content (or about motivating subjects to use predicate logic)

fails, because subjects are producing reasoning patterns strikingly at variance with predicate logic, which are consistent with the logic of social exchange.

Fodor's Final Challenge

As discussed, Fodor believes (incorrectly) that social contract theory predicts reasoning performance will be insensitive to logical form. He challenges us to pit our two views against each other by showing that varying logical form has no effect when social contract content is held constant.

However, social contract theory (and other, comparable theories about evolved specializations such as hazard management theory) predict the opposite: that inference will be sensitive to logical form (under the relevant conditions). Moreover, these theories pose deep and interesting questions about what *counts* as a difference in logical form *as defined by the architecture of the human mind*. Two deontic conditionals that Fodor interprets as "having" the logical form "it is required that (If *P* then *Q*)" should be assigned different logical forms by social contract algorithms when one fits the benefit-requirement syntax of a social contract and the other does not.

Nevertheless, social contract theory does not predict that logical form (in Fodor's sense) will have an effect regardless of any other factor. It pinpoints conditions under which the effects of logical form will appear and disappear. This makes it possible to meet Fodor's challenge.

Indeed, from our earliest experimental publications (e.g., Cosmides, 1989), we have produced a series of cases in which we did what Fodor claims is necessary. That is, we held social contract content constant, varied logical form, and showed that this has no effect on choosing the *requirement not met* card. For example, in experiments comparing standard and switched social contracts (section IV.iii.i), social contract content and social contract logical form were both held constant. What varied was logical form *as defined by predicate and deontic logics*, from "If you take the benefit, then you (must) satisfy the requirement" to "If you satisfy the requirement, then you (may) take the benefit."⁷ Figure 2.4.1 shows that the percent of subjects choosing the "requirement not met" card is invariant over this change in logical form, both at Harvard and in the Ecuadorian Amazon. In a more subtle way, the perspective change experiments are also relevant: the "requirement not met" card was chosen as often for agent 1 as for agent 2, whether this corresponded to the *not-Q* card or the *not-P* card. The rule given only specified a requirement for agent 1 (e.g.,

employees). To detect cheating by agent 2 (the employer), a reciprocal obligation that falls on agent 2 had to be inferentially derived from the rule given, and this entails a (covert) transformation of logical form (see section IV.iii.ii for an explanation).

In sum, Fodor's alternative explanation for the performance effects triggered by manipulating social exchange variables has been falsified in a large number of independent ways. It is time that Fodor stop repeating the factual error that the primary evidence for social contract theory is a difference between reasoning on indicative and deontic conditionals and begin to grapple with the far broader and more challenging array of results testing this theory. Missing from our discussion is what is most distinctive about social contract theory and hazard management theory: that a large series of precise, surprising, and confirmed predictions about how people reason were derived in a principled way from a careful consideration of selection pressures acting on our ancestors. Readers who want to explore this larger array of findings could begin with the Cosmides and Tooby paper in this volume that sparked this exchange with Fodor. Philosophers might want to consider how the details of this dispute provide a case study of why the possible existence of content-sensitive reasoning systems in the human mind call into question (or at least bring into scrutiny) the deepest feature of traditional logics, their sharp distinction between form and content.

Notes

1. In his reply, Fodor justifies his obliviousness of the primary empirical literature that he claims to have explained by saying that Buller (2005) has reviewed it and agrees with him that there are systematic confounds between logical form and content. Fodor really needs to read these things himself, rather than—in true blind leading the blind fashion—relying on Buller's (2005) account. After all, Buller derives his 2005 account from Fodor (2000), and his "review" systematically ignores large numbers of experiments that falsified his and Fodor's view. For example, one would not know from reading Buller that there are any of the many experiments that hold deontic logical form constant, comparing deontic conditionals with and without social contract content. Bizarrely, Fodor describes us as "curiously reticent" about replying to Buller's confound charges when (among other papers) the paper Fodor is commenting on is a reply: we went into tedious detail on the various experiments holding logical form constant while manipulating key social contract variables. We also went into tedious detail explaining that social contract theory posits domain-specialized interpretive rules, and what they are. This was because Buller's critique posits that people make precisely the inferences that social contract theory predicts

they will—that is, these inferences are just post hoc restatements of the predictions that social contract theory makes (but without providing an alternative causal mechanism).

2. Indeed, Fodor seems confused when he talks about Beaman's results. *Fodor* is the one proposing that high levels of cheater detection reflect nothing more than a deontic/indicative manipulation, yet he says, "Beaman finds the effect of the deontic conditional/descriptive conditional manipulation on performance on the Wason task *is very significantly reduced* when the confound is removed" (italics ours; p. 139). That is, when content (vowel vs. drinking age) is held constant, the effect of Fodor's logical form variable is too small to account for social contract performance. But when logical form is held constant, Beaman found that the effect of the presence/absence of social contract content was very large in comparison.

3. It uses a "must" rather than "it is required that," but that shouldn't bother Fodor. Until Beaman (2002), no one had tested deontic social contracts using Fodor's requirement language (and no one had tested his "under 18→coke" version of the drinking age law). Thus, all the data Fodor was trying to explain in 2000—all the evidence of high-performing social contracts—used "must" or no deontic operator, yet Fodor considered that his proposal equally well accounted for results on those problems.

4. Of course, this rule does not really require you to be over 18. Pace Fodor, it requires you be over 18 *when* you drink beer *and* there is a drinking age law—something that is perfectly possible for anyone to abide by.

5. Indeed, as the linchpin of his account of Wason task results, Fodor made an appeal to intuition, claiming that the contrapositive of an indicative conditional is not parallel to the contrapositive of a deontic conditional. When we pointed out that this claim was considered debatable by logicians, Fodor resorted to ridicule—a form of argument from authority. Fodor suggested that the only philosopher who could have made such a contra-Fodorian argument was someone so hapless, he or she should never have been tenured. The proposition Fodor resists is that indicative conditionals can be seen as having a logical form parallel to that which he claims for deontic ones, that is, as asserting Q in the case that P . In other words, Fodor disputes that *if P then Q* "is commonly felt less as an affirmation of the conditional than as a conditional affirmation of the consequent." On principle, we resist arguments from authority, preferring to deal with the merits of an issue. However, for those who believe in the division of intellectual labor and want to know the relative authorities of the disputants, we point out that this quote is from a famous logic textbook by Willard Van Orman Quine (1972; see p. 19 for more explanation), who attributes the insight originally to Philip Rhinelander. Although Fodor (relying once again on his intuition) may not consider either Quine or Rhinelander his equals as logicians, we can at least testify that both got tenure—at Harvard and Stanford, respectively. In any case, Quine's interpretation is a way of capturing the

(widespread) intuition that cases of *not-P* don't really confirm $P \rightarrow Q$ as a whole (e.g., are black crows really evidence confirming "If x is a swan, then x is white," or are crows merely irrelevant in evaluating a conditional about swans?). It is worth pointing out that this interpretation is borne out by considering the evolution and evolved function of conditional reasoning and the functional regulation of its scope (Cosmides & Tooby, 2000b). From an adaptationist perspective, the conditional is designed to be activated by the input conditions specified in the antecedent and is simply irrelevant (computationally inactivated) when the antecedent condition is not met (in reality, or suppositionally). This would also explain why *modus ponens* is easily activated in subjects, while *modus tollens* is not spontaneously deployed: not-white is computationally not an input condition that activates the conditional, "If x is a swan, then x is white."

6. Fodor says it is not naive realism to say a sentence "has" an interpretation, because he really meant "has relative to a grammar L." But Fodor continues to presume that the mind has a single logical grammar. If there are many different grammars in the mind—different inference systems, each implementing a different syntax—then the mind will assign different logical forms to a conditional depending on whether the inferential system doing the interpreting is predicate logic, social contract algorithms, the hazard management system, a domain-general deontic logic (if such exists), and so forth. Determining which logical form a conditional "has" then becomes a matter of (theoretically guided) empirical discovery. The issue cannot be decided by a priori arguments like Fodor's, which presuppose there is only one kind of logical form. That is why a central goal of our research has been to test whether the different syntaxes of each proposed evolved logic predict which content manipulations will trigger changes in representational format and reasoning outcome—e.g., what content manipulations change a deontic rule into a social contract, and what distinguishes a social contract from a precautionary rule. *How* the mind represents different conditionals is precisely what is under investigation.

7. "Must's" and "may's" were sometimes present, and sometimes they were absent—this didn't change the results. For example, the switched social contract in the Amazon had a "may," and the switched ones in Cosmides (1989) did not—both are in figure 2.4.1.