Essays

Rationality and Counterfactual Legal Analysis

Antonio Estella de Noriega*

Abstract

The aim of this article is to argue that counterfactual legal analysis should be used as a primary method in judicial interpretation of legislation. The article examines this issue assuming a rationality setting in which law is understood as a credibility device. Judges should show deference to the legislator when counterfactuals have been foreseen by the latter; in contrast, they might substitute their own judgment for the legislator's choices when the latter has not devised a counterfactual situation, and the norm is not an equilibrium from a 'law as credibility' perspective.

'Bad promises are better broken than kept'.

Abraham Lincoln
April 11, 1865.[1]

I. Introduction

We often rely on counterfactual thinking, as in 'if I had married Tony instead of Philip, I would have been happier'. The famous Woody Allen counterfactual statement (had I met God, I would have told him not that he is evil – just that he is an underachiever) is but another proof of this. In more general terms, the arts, like movies, or novels, or paintings, can be seen as fictional counterfactuals – as stories of what could have happened if we had been different people. We sometimes want to be those people – if only we could live those alternative stories.[2]

* Associate Professor of Administrative Law and Jean Monnet Professor 'ad personam' of European Economic Governance Law, University Carlos III of Madrid.

[1] Collected works of Abraham Lincoln available at http://quod.lib.umich.edu/l/lincoln/lincoln8/1:850?rgn=div1;singlegenre=All;sort=occur;subview=detail;type=simple;view=fulltext;q1=april+11+1865 (last visited 24 May 2016).

[2] For example, as for novels, this is precisely the point that is made by J. Cercas, *El punto ciego: Las conferencias Weidenfeld 2015* (Barcelona: Penguin Random House Grupo Editorial España, 2016). Cercas argues: '*Es mentira (…) que las novelas sirvan solo para*

Counterfactual thinking is a rather sophisticated kind of thinking. It requires a certain level of psychological and neuronal development. In an interesting article on the relationship between law and the brain, Abigail A. Baird and Jonathan Fugelsang[3] give evidence of the (to a certain extent) counterintuitive finding that adolescents are less capable of counterfactual thinking than adults, due to their relatively less mature brains.[4] Counterfactual thinking is therefore complex.

That we use counterfactual thinking all the time (to surmise what could have been if we had done something different) and that this kind of thinking is demanding, does not necessarily provide for a good argument in favour of its being used in social sciences; however, as we know, counterfactual thinking *is* used in the social sciences. History – as an academic discipline – is a prime example of this. In fact, it can be argued that serious counterfactual thinking is born in history.[5] However, it is not restricted to this discipline: political science, economics, sociology and even psychology, all rely on counterfactual thinking. But the question is: what about law?

The idea that counterfactual thinking must be used, and in fact is used, in the legal realm is not entirely new. However, it is rather restricted to what is called the 'evidence-finding-and-giving' phase of judicial activity. The aim of this contribution is to argue that counterfactual thinking is also fit for legal interpretative purposes. Not only this – the aim of this article is to show that counterfactual thinking should be a primary method for approaching judicial interpretation of legislation.[6] The question that this paper will address is therefore the following: can counterfactual legal interpretation be enlarged beyond the evidence-finding-and-giving phase of the judicial process? If so, how can this be done?

To this end, I shall proceed as follows. In the next section, Section 2, I

pasar el rato, para matar el tiempo; al contrario: sirven (...) para hacernos como nunca hemos sido'. ('It is untrue that novels only serve the purpose of having a good time; on the contrary, they serve the purpose of making us as we have never ever been').

[3] A. Baird and J. Fugelsang, 'The Emergence of Consequential Thought: Evidence from Neuroscience', in S. Zeki and O. Goodenough eds, *Law and the Brain* (Oxford-New York: Oxford University Press, 2004), 245-258.

[4] This is counterintuitive since one would tend to think that adolescents are more prone than adults to delusional thinking, to imagining parallel and fictional worlds. Fictional worlds are not counterfactual, but they are made of them.

[5] See P. Tetlock and A. Belkin, 'Counterfactual Thought Experiments in World Politics: Logical, Methodological and Psychological Perspectives', in P. Tetlock and A. Belkin eds, *Counterfactual Thought Experiments in World Politics. Logical, Methodological and Psychological Perspectives* (Princeton: Princeton University Press, 1996), 1-38, 3: 'There is nothing new about counterfactual inference. Historians have been doing it for at least two thousand years'.

[6] This could be applied as well to private contracts, but I shall not deal with this point in this article.

will show the basic structure of counterfactual thinking, in general terms. In Section 3, I will show and discuss a number of examples in which counterfactual thinking is used by the courts. Sections 4 and 5 will deal with the question of how counterfactual thinking can be extended from the evidence-giving phase of a trial to the analysis of the equilibrium path of a given norm. Finally, the concluding section of this article, Section 6, will provide a very simplified version of this paper's argument as well as some final remarks.

II. The Structure of Counterfactual Reasoning

Let us now fine-tune what we may understand by counterfactual reasoning. The first point to be made is that counterfactual reasoning has a link with facts. This is why it is of no surprise that where this kind of reasoning has made most headway has been in history as an academic discipline. Therefore, the first thing that we need to have is a certain set of facts which are proved, or deemed to be true. I shall refer to them as 'F'.

The second thing that is needed to make sound counterfactual reasoning is a consequence. Therefore, given F, a particular consequence is produced. I shall refer to this consequence as 'C'.

If we tie together the two previous elements, what we have is the following structure:

given F, then C

This is where counterfactual reasoning operates, or impacts upon. To start with, the question that counterfactual reasoning tries to solve is the following: how do we know that F produced C? How can we be sure of that? Though a truism, it is worthwhile remembering that we are speaking here not of natural sciences or physics but *of the course of human events*. Therefore, the ultimate proof that F produced C in social sciences is very difficult to obtain, if not, in many cases, impossible. This is even more so the case in situations of multi-causality, where there are multiple causes that could have produced a particular consequence. What we want to isolate is which cause, or which causes, *directly* produced C.

Therefore, one of the ways that we can ascertain whether F produced C is to ask about the alternative courses of actions that could have produced C. This is what we call counterfactual reasoning. Counterfactual reasoning is therefore a thought experiment in which we inquire about the alternative factual cause or causes that could have produced a given outcome. Thus, the structure of a counterfactual thought experiment is the following: if F1 had happened, would C have happened as well? If the

answer we give to this question is positive, then F is false. On the contrary, if the answer we give to this question is negative, then F would be true.

Let us give an example to shed some light on what can be seen as a very obscure and cold discussion. A very good example is provided by the decision taken by the US government, at the end of the Second World War, to drop two atomic bombs on Japanese soil. This has been deemed to be one of the most contested decisions ever. In fact, it spurred a social movement in the US – and beyond – of anti-war and anti-atomic energy sentiments, among other things. The main way of justifying this decision by the US government was to say that if the atomic bombs had not been dropped on Hiroshima and Nagasaki, then the war would have ended much later, and the toll of lives (of American lives, of course, but also of other countries' lives) would have been much higher. In other words, the costs of dropping the bombs were much lower than the costs of not doing so, and the benefits of doing so much higher than the benefits of not doing so.

However, this is, as is known, very controversial. How do we know this to be true? How do we know that if the bombs had not been dropped, the imperial army would not have been defeated in a question of months? Ulterior research on the issue[7] shows that in fact the Japanese army was already very weak when the bombs were dropped, and that there were movements within the Japanese government (not the least important those of the Emperor Hirohito himself)[8] to stop the war as soon as possible. Maybe, if the bombs had not been dropped, the outcome would have been

[7] See, in particular, G. Gentile, *How effective is Strategic Bombing? Lessons learned from World War II to Kosovo* (New York-London: The New York University Press, 2001).

[8] See the United States Strategic Bombing Survey Summary Report (Pacific War) of 1 July 1946, available at http://www.anesi.com/ussbs01.htm#teotab (last visited 24 May 2016). This is the Report on Strategic Bombing in Japan that Paul Nitze (director and then as Vice Chairman of the Strategic Bombing Survey) presented to the US Senate Committee on Atomic Energy, after the Second World War. The report is, precisely, a fascinating counterfactual-policy exercise, as shall be shown in the next footnote. As regards the Emperor Hirohito's role in shortening the war, the report says on page 26: 'On 20 June (1945) the Emperor, on his own initiative, called the six members of the Supreme War Direction Council to a conference and said it was necessary to have a plan to close the war at once, as well as a plan to defend the home islands. The timing of the Potsdam Conference interfered with a plan to send Prince Konoye to Moscow as a special emissary with instructions from the cabinet to negotiate for peace on terms less than unconditional surrender, but with private instructions from the Emperor to secure peace at any price. Although the Supreme War Direction Council, in its deliberations on the Potsdam Declaration, was agreed on the advisability of ending the war, three of its members, the Prime Minister, the Foreign Minister and the Navy Minister, were prepared to accept unconditional surrender, while the other three, the Army Minister,

very similar to that obtained with the dropping of the bombs, since, for example, Japan was exhausted, both psychologically and economically, as a result of the previous war years. Due to exhaustion (F1), the war would have ended anyway (C). If this were true, it would render F false, or at least, revisable.[9]

At this point it is necessary to introduce some restrictions to the way in which we carry out counterfactual experiments. This point has been raised by Jon Elster and it is worth remembering here. Jon Elster[10] argues that counterfactual reasoning has to be plausible. According to Jon Elster, plausibility means, at the very least, internal consistency between the antecedent and the consequent, and historical possibility.[11] As regards the first – internal consistency – the rules of logical reasoning apply here. The second aspect – historical possibility – is more important, as it implies that the counterfactual story has to be as close as possible to the factual one. If we said that Japan would have surrendered and the Second World War ended had Martians attacked Earth, this might be true, but it is not very plausible as a counterfactual. To do this, we would first and foremost have to ascertain that there is life on Mars; secondly, that the Martians are developed enough as to make quick trips to Earth; thirdly, that the Martians have an army; fourthly, that they have motives to attack Earth; and so on and so forth. In contrast, the counterfactual, 'had the Americans not dropped the bombs, but just kept engaging in conventional warfare, the war would have also ended in a question of months', seems much more plausible. The Americans had the capacity to do this and it also seems that

and the Chiefs of Staff of both services, favored continued resistance unless certain mitigating conditions were obtained'.

[9] This is made clear in the Report mentioned in n 8 above. On page 26, it is said: 'Based on a detailed investigation of all the facts, and supported by the testimony of the surviving Japanese leaders involved, it is the Survey's opinion that certainly prior to 31 December 1945, and in all probability prior to 1 November 1945, Japan would have surrendered even if the atomic bombs had not been dropped, even if Russia had not entered the war, and even if no invasion had been planned or contemplated'.

[10] J. Elster, *Making Sense of Marx* (Cambridge: Cambridge University Press, 1985).

[11] P. Tetlock and A. Belkin, n 5 above, 18, add to internal consistency and historical possibility (or to what they call the 'minimal re-write rule') the following four conditions: clarity, theoretical consistency, statistical consistency and projectability. As for clarity, this raises a more general point on methodology: this article belongs to the analytical tradition, which has clarity as one of its main pillars, as is known. Therefore it would be redundant, in my opinion, to add this condition. As regards theoretical consistency, I think that it could be subsumed into the first condition (logical or internal consistency). As for statistical consistency, it is irrelevant, in principle, for our discussion on legal counterfactuals. And as regards projectability, it is difficult to see the difference between this condition and plausibility or the minimal rewrite rule, since historical consistency means that the new facts (F1) have to have some connection with reality (in fact, the larger the link, the better).

one of the plans they were considering was precisely to keep engaging in conventional warfare against Japan.[12]

III. Counterfactual Thinking in Law: The US Jurisprudence on Habeas Corpus

Let us turn now to the use of counterfactuals in the legal world. As is known, counterfactual reasoning is currently being used by the US courts. More generally, the US legal context is the legal tradition in which the debate around the use of counterfactual reasoning in legal interpretation has made most headway.[13] In this section, I shall simply show how US courts, and in particular, the US Supreme Court, is making use of counterfactual reasoning, through the analysis of three hallmark cases on habeas corpus. These cases are the following:

- *Harrington, Warden* v *Richter* (19 January 2011)
- *Cullen, Acting Warden* v *Pinholster* (4 April 2011)
- *Greene, Aka Trice* v *Fisher, Superintendent, State Correctional Institution at Smithfield, et al* (8 November 2011)

I shall review these three cases below. But before I do so, it is important to make two points. First, all these cases originate in the context of the so-called US 'habeas corpus' law procedure, which implies a revision by a different court of the sentence that another court gave on a particular penal case (most of which involve the death penalty). Second, the applicable law to these cases is the AEPDA, the Antiterrorism and Effective Death Penalty Act of 1996, which modifies the previous legislation on habeas corpus. In particular, Art 2254 d) of title 28 of the United States Code under AEPDA states that habeas corpus shall not be granted unless the previous sentence:

> '(1) resulted in a decision that was contrary to, or involved an unreasonable application of, clearly established Federal law, as determined by the Supreme Court of the United States;

[12] See, again, United States Strategic Bombing Survey Summary Report (Pacific War) of 1 July 1946, n 8 above.

[13] See, in particular, A. Burns, 'Counterfactual Contradictions: Interpretative Error in the Analysis of AEDPA' 65 *Stanford Law Review*, 203-239 (2013); R. Strassfeld, 1992, 'If...Counterfactuals in the Law', Case Western Reserve University, School of Law, Faculty Publications, Paper 373. The discussion is though not limited to the US context: for an example in the Dutch legal context, see N. Nivelle, 'Counterfactual Conditionals in Argumentative Legal Language in Dutch' 18(3) *Pragmatics*, 469-490 (2008).

or

(2) resulted in a decision that was based on an unreasonable determination of the facts in light of the evidence presented in the State court proceeding'.

Therefore, counterfactual arguments are very much connected to the '(un)reasonableness' clause included therein.[14]

1. *Harrington, Warden* v *Richter* (19 January 2011)

The facts of this case are the following. A shooting occurred in the domicile of a drug dealer by the name of Johnson. According to Johnson's account, Richter and Branscombe (with whom Johnson had been smoking marijuana some hours before) shot Johnson and his colleague Klein. According to Richter's account, he was not involved in the shooting. However, once he was arrested, he later admitted that he had been involved in the shooting but only to dispose of the guns that Branscombe had used to shoot Johnson and Klein. Richter's lawyer did not present expert testimony on serology. Richter was found guilty in first instance and sentenced to life without parole. On appeal, his conviction was confirmed. Furthermore, the California Supreme Court denied his petition for review. Richter later petitioned the California Supreme Court for a writ of habeas corpus. He argued that his lawyers should have presented expert testimony on serology and that this caveat affected the final outcome. His was a classical counterfactual reasoning: had his lawyers presented serology evidence, he would have been found innocent of the murder and the sentence mitigated. Habeas corpus writ was initially denied, but the Court of Appeals granted relief. Finally, the Supreme Court of the United States reversed the Court of Appeals decision and denied the writ of habeas corpus.

The US Supreme Court used a counterfactual-like reasoning at some points in its judgement. A first clear example of this would be the following:

'With respect to prejudice – says the Court – a challenger must demonstrate "a reasonable probability that, but for counsel's unprofessional errors, the result of the proceeding would have been

[14] There is already abundant literature on reasonableness and the law. See: P. Craig, 'The Nature of Reasonableness Review' 1 *Current Legal Problems*, 1-37 (2013); C. Sunstein, 'Cost-Benefit Analysis Without Analyzing Costs or Benefits: Reasonable Accommodation, Balancing, and Stigmatic Harms' 74 *University Chicago Law Review*, 1895-1909 (2007); C. Sunstein, 'The Myth of the Balanced Court' 18(9) *The American Prospect*, 28-29 (2007). For the Italian legal context, see G. Perlingieri, *Profili Applicativi della ragionevolezza nel diritto civile* (Napoli: Edizioni Scientifiche Italiane, 2015).

> different. A reasonable probability is a probability sufficient to undermine confidence in the outcome". It is not enough "to show that the errors had some conceivable effect on the outcome of the proceeding". Counsel's errors must be "so serious as to deprive the defendant of a fair trial, a trial whose result is reliable" '. (page 15, syllabus)[15]

In this paragraph, the Court uses the well-known, in US law, 'but-for' clause. In my opinion, the 'but-for' clause is the closest expression to a traditional counterfactual-like kind of reasoning in law: had it not been for this, then the outcome would have been different. In *Richter*, this amounts to arguing that had the lawyer provided serology evidence, then Richter's line of argument would have been proven, and therefore he would not have been found guilty of Klein's murder. However, in this case, what the Court is saying is that the challenger has to 'demonstrate' *with a high degree of probability* that this would have been the case; in other words, the challenger should demonstrate that the counterfactual is plausible. How this should be done is something that the Court does not say. But what is clear is that we are here in the world of evidence: the challenger would have to provide evidence that the counterfactual is true or at least plausible.

A second example would be the following:

> 'The Court of Appeals opinion for the en banc majority rests in large part on a hypothesis that reasonably could have been rejected. The hypothesis is that without jeopardizing Richter's defense, an expert could have testified that the blood in Johnson's doorway could not have come from Johnson and could have come from Klein, thus suggesting that Richter's version of the shooting was correct and Johnson's a fabrication. This theory overlooks the fact that concentrating on the blood pool carried its own serious risks. If serological analysis or other forensic evidence demonstrated that the blood came from Johnson alone, Richter's story would be exposed as an invention. An attorney need not pursue an investigation that would be fruitless, much less one that might be harmful to the defense. (Strickland, supra, at 691). Here Richter's attorney had reason to question the truth of his

[15] According to the Court, this probability was not demonstrated by the habeas corpus petitioner. A similar line of argument is used here: 'With respect to defense counsel's performance, the Court of Appeals held that because Richter's attorney had not consulted forensic blood experts or introduced expert evidence, the California Supreme Court could not reasonably have concluded counsel provided adequate representation. This conclusion was erroneous' (page 16, syllabus).

> client's account, given, for instance, Richter's initial denial of involvement and the subsequent production of Johnson's missing pistol'. (page 18, syllabus)[16]

In this paragraph, the counterfactual is also very clearly made by the Court. For the Court, in effect, the argument is that had serology evidence been given, then the outcome could have been different. However – and this is the point for the Court here – what this argument fails to see is that pursuing this strategy also had its own risks: basically, to reach the contrary outcome to the one Richter was arguing (that the blood came from Johnson alone and not from both Johnson and Klein). If this had been the case, then Richter's argument would have been destroyed. This is an interesting case of a counterfactual, since what the Court is doing here is to destroy the first counterfactual (had serology been done, then the outcome would have been different) with a second one (but what if serology had been done, with the outcome being the same). The second counterfactual helps the Court to identify the risks implicit in the strategy of using serology evidence. In principle, this counterfactual is a real thought experiment, which is not linked with giving further evidence or not. However, the Court does not say anything about the question of how real that risk – the risk that the serology evidence would have destroyed Richter's main line of argument – was, which clearly remits us to an evidence-giving (at least in probabilistic terms, as in the first paragraph) problem.

Finally, the partially dissenting opinion of Justice Ginsburg is interesting also as regards the use of counterfactuals. In Ginsburg's opinion:

> 'In failing even to consult blood experts in preparation for the murder trial, Richter's counsel, I agree with the Court of Appeals, "was not functioning as the 'counsel' guaranteed the defendant by the Sixth Amendment." Strickland v. Washington, 466 U. S. 668, 687 (1984). The strong force of the prosecution's case, however, was not significantly reduced by the affidavits offered in support of Richter's habeas

[16] In the same vein: 'Under §2254(d), a habeas court must determine what arguments or theories supported or, as here, could have supported, the state court's decision; and then it must ask whether it is possible fair minded jurists could disagree that those arguments or theories are inconsistent with the holding in a prior decision of this Court. The opinion of the Court of Appeals all but ignored "the only question that matters under §2254(d)(1)". Lockyer v. Andrade, 538 U.S. 63, 71 (2003)' (page 12, syllabus). And again: 'The Court of Appeals appears to have treated the unreasonableness question as a test of its confidence in the result it would reach under de novo review: Because the Court of Appeals had little doubt that Richter's Strickland claim had merit, the Court of Appeals concluded the state court must have been unreasonable in rejecting it' (page 12, syllabus).

> petition. I would therefore not rank counsel's lapse "so serious as to deprive [Richter] of a fair trial, a trial whose result is reliable." For that reason, I concur in the Court's judgment'.

This counterfactual is interesting since what Ginsburg is saying is that if serology tests had been done, then the outcome would have been more or less the same. Thus she agrees with the Court of Appeals that serology is a test that, in this case, should have been done – which proves a bad counselling praxis – but that even if it had been done, the outcome would not have been so decisive as to challenge the conclusion, which is that Richter was guilty of murder. Again this remits us to an evidence-finding problem.

2. *Cullen, Acting Warden* v *Pinholster* (4 April 2011)

The facts of this case are the following. Scott Lynn Pinholster, together with two other accomplices, broke into a house with the purpose of committing burglary and brutally killed two men who interrupted the robbery. A jury found Pinholster guilty and sentenced him to death**.** The California Supreme Court (twice) denied habeas corpus to Pinholster, but the federal District Court reversed and granted relief. The Court of Appeals of the Ninth Circuit affirmed. Pinholster argued that his lawyers had failed to 'adequately investigate and present mitigating evidence, including evidence on mental disorders'. For example, a psychiatrist diagnosed Pinholster with bipolar mood and seizure disorders. Had this been done, the outcome would have been different, he argued. However, the Supreme Court reversed. Justice Sotomayor gave a long dissenting opinion in this case.

The first example of the use by the Court of a counterfactual is the following:

> 'The Court also required that defendants prove prejudice. "The defendant must show that there is a reasonable probability that, but for counsel's unprofessional errors, the result of the proceeding would have been different." "A reasonable probability is a probability sufficient to undermine confidence in the outcome." Ibid. That requires a "substantial," not just "conceivable," likelihood of a different result. Richter, 562 (…)'. (page 17, syllabus)

Again, the Court uses here the 'but-for' clause, much in the same vein as it did in *Richter*. As was said before, the 'but-for' clause is the closest we have to classical counterfactual reasoning.

A second example is this one:

> 'There is no reasonable probability that the additional evidence Pinholster presented in his state habeas proceedings would have changed the jury's verdict'. (page 29, syllabus)

Here the Court is saying that the additional evidence that Pinholster gave in the case (testimonies of his psychologically undermined condition, due to a de-structured infancy in which he was brutalised by his father, and similar episodes) does not amount to affecting the outcome with a 'high probability'. Again, the Court fails to define what a high probability would be (reasonableness test), but this clearly remains within the realm of evidence-finding and -giving.

The third example is given by Justice Sotomayor's dissenting opinion. Her lengthy opinion is full of counterfactual reasoning, but the major point that she makes is well summarised by the following:

> 'In sum, the evidence confirmed what was already apparent from the state-court record: Pinholster's counsel failed to conduct an adequate mitigation investigation, and there was a reasonable probability that at least one juror confronted with the "voluminous" mitigating evidence counsel should have discovered would have voted to spare Pinholster's life. Ibid. Accordingly, whether on the basis of the state- or federal-court record, the courts below correctly concluded that Pinholster had shown that the California Supreme Court's decision reflected an unreasonable application of Strickland'. (page 42, syllabus)

The interesting aspect of Sotomayor's counterfactual reasoning is that through this approach, she gets to exactly the opposite conclusion than the majority of the Supreme Court. In her opinion, the evidence that Pinholster gave at a later stage of the process did constitute mitigating evidence. If this evidence had been given at the beginning of the process by his counsel, there would have been, according to Sotomayor, a 'reasonable' probability that at least one juror would have voted against applying capital punishment on Pinholster. Again, we do not know the factual basis of this argument, but it seems of a rather empirical kind: it would seem as if Sotomayor had, for example, statistical evidence that when certain evidence (evidence on mental disorders) is given, then jury unanimity in applying capital punishment in a murder case would fall apart. We still remain in the realm of evidence-giving and -finding here.

3. *Greene, Aka Trice* v *Fisher, Superintendent, State Correctional Institution at Smithfield, et al* (8 November 2011)

The facts of this case are the following. Eric Greene, together with four accomplices, robbed a grocery store in North Philadelphia (Pennsylvania). During the robbery, one of the men shot and killed the grocery store's owner. A jury convicted Greene of murder. He appealed to the Pennsylvania Supreme Court, which dismissed the application. Green then filed a writ for habeas corpus at the District Court for the Eastern District of Pennsylvania, alleging, inter alia, that the introduction of his non-testifying co-defendants' statements violated the Confrontation Clause. Had they not confessed, the outcome would have been different, according to Greene. The District Court denied the petition. However, the Court of Appeals granted relief. The US Supreme Court then reversed.

The clearest example of a counterfactual reasoning being used in this case is the following:

> 'Where intervening developments ... reveal a reasonable probability that the decision below rests upon a premise that the lower court would reject if given the opportunity for further consideration, and where it appears that such a redetermination may determine the ultimate outcome of the litigation, [an order granting the petition, vacating the judgment below, and remanding the case (GVR)] is, we believe, potentially appropriate'. Lawrence v. Chater, 516 U. S. 163, 167 (1996) (per curiam) (page 7)

In other words, what the Court is using here is a very classical argument in counterfactual terms. This argument is as follows: if there were new evidence which, with a high probability, could alter the ultimate decision in a given litigation, then this evidence should be taken into account. Once again, evidence is the key issue here.

In conclusion, all this shows, firstly, that the Supreme Court regularly uses counterfactual analysis in its rulings. The second point the previous analysis shows is that the issue of counterfactual reasoning is rather connected to the fact-finding or evidence-giving part of the trial. In effect, all of these cases are cases in which the question posed is whether new or different or further evidence should have been given in a particular process. The Court has to be persuaded that had this evidence been given, then the outcome would had been altered with a high probability (that is, then the outcome could have been different to a reasonable scale). When the Court is convinced of this point, then it grants habeas corpus, and vice-versa. The counterfactual reasoning is therefore key for the final decision that the Supreme Court reaches, but on the other hand it is always directly linked to the evidentiary phase of a given trial.

IV. Extending the Use of Counterfactuals: Rationality and Law as Credibility

My aim in the remaining sections of this article is to extend the use of counterfactual legal reasoning from the realm of trial evidence to the realm of the factual basis upon which norms rest. That is, what I am proposing is to enlarge the use of counterfactuals. The argument is that counterfactual reasoning should become a primary legal technique that should be used in legal interpretation.

To bring the argument home, I first of all need to make two preliminary points: one on what I understand by rationality and the law, and the other on what I call 'law as credibility'. Then, I will move on to the question of how to enlarge counterfactual analyses of law.

1. Rationality and the Law

When I refer to rationality and the law I am not referring to the usual sense in which rationality is used in legal jargon. For example, Paul Craig[17] analyses how the principle of rationality is being used by the English courts to control for administrative decisions. In turn, Patrick Birkinshaw[18] traces the roots of rationality to other principles and legal concepts that are employed in different legal traditions than the English one. Their common point is that a rationality check looks at the intrinsic logic of the reasoning that a given agency is making to support a particular decision that it has taken. If the courts unearth an illogical step in the agency's reasoning, they will annul the decision, but they will preserve it if it is logical.[19]

My approach is rather different. When I speak of rationality and the law, I am only setting the analytical framework, or the premise, of a much wider debate in which the idea of law as credibility can be better understood.

[17] P. Craig, 'Proportionality, Rationality and Review' *New Zealand Law Review*, 265-301 (2010).

[18] P. Birkinshaw, *European Public Law: The Achievement and the Challenge* (Alphen aan den Rijn: Wolters Kluwer, 2nd ed, 2014). See A. Estella de Noriega, 'Book review: *European Public Law: The Achievement and the Challenge*, Second Edition, by Patrick Birkinshaw' 22(2) *European Public Law*, 397-404 (2016).

[19] Paul Craig cites Lord Diplock here: rationality review is limited to a 'decision which is so outrageous in its defiance of logic or of accepted moral standards that no sensible person who had applied his mind to the question could have arrived at it' (*Council of Civil Service Unions* v *Minister for the Civil Service* [1985] AC 374 at 410). Though the author argues in favour of a broader understanding of rationality, I think that the only way to differentiate between a rationality review and a reasonableness review is that the first one looks at the intrinsic logic of the reasoning that an agency gives to adopt a decision, while the second one, more broadly, relies on a balancing exercise of (to a certain extent) contradictory aims.

If we do not clarify what this framework is, then the idea of law as credibility is difficult to understand. Therefore, the framework I envision is the following.

The basis of my analysis stems from rational choice and game theory. As is known, for these intellectual traditions, the basic unit of analysis is the individual. This basic framework is then applied to collective agents, such as institutions. The legislator, the executive, the courts, all of them, are taken as if they were one individual. The main reason for this exercise of analytical reductionism is that it allows the attribution of preferences to these actors and also the capacity to develop strategic plans to attain such preferences. It also allows attributing motivations to institutions, which is the key point in this discussion.

Therefore, institutions, like individuals, have preferences, develop strategies to attain those preferences, and have motives to act. In my particular framework, a number of important points have to be made, which are all of relevance for a legal discussion. First, as for preferences, I understand, as in Marx,[20] that they are exogenously induced.[21] This means that institutions, like individuals, do not set their utility functions in complete freedom, but that they are very much constrained by exogenous forces to set up their utility functions. If we prefer a less ambiguous word, individuals' preferences are pre-determined, to a large extent at least. The second point that I wish to make here is that individuals cannot be deemed to be super-humanly clever in enacting strategies: they have a limited capacity to devise and build strategies. And above all, they suffer from many problems of information, which make their strategies very imperfect and rather simple in the majority of cases. The third point regards motives. In a very schematic version of rational choice, the basic individual's motive is egoism. However, we know from game theory experiments in laboratories and from behavioural economics that the truth is somewhat more complex: individuals are not only driven by sheer egoism. In my opinion, they are driven by self-interest – which is a much more complex notion. I understand self-interest as the production of a feeling of warm-glow, of internal welfare. Individuals attempt to increase this feeling as much as they can, and this can explain, for example, altruistic behaviour.

Let us apply this framework to law. The first point to be mentioned is that I see the relationship between institutions and the people as a game

[20] K. Marx, *The eighteenth Brumaire of Louis Bonaparte* (London: George Allen & Unwin, 1926), translated from German by P. Eden and P. Cedar.

[21] 'Man makes his own history, but he does not make it out of the whole cloth; he does not make it out of conditions chosen by himself, but out of such as he finds close at hand. The tradition of all past generations weighs like an alp upon the brain of the living': K. Marx, n 20 above, chapter I.

in which both have particular preferences, strategies, and motives. As in Hobbes,[22] people establish institutions to overcome their state of nature. However, as in Rousseau,[23] once they establish institutions, they start mistrusting them. They are not sure that institutions will attend to the objectives for which they were set up (which is basically to help to reduce the problems that collective action involves). Here is an example: in a state of nature which is dominated by violence, the people decide to set up institutions which have the monopoly of violence. They decide to withdraw all weapons from the society and give them to the institutions, so that people will not kill each other anymore. However, the institutions have their own preferences: for example, to become wealthy. They do this because it makes them feel well. And they use strategies to attain this objective. Thus, institutions, once the guardians of peace, use the weapons that the society gave them to dominate the people with the purpose of becoming wealthy. As the people know this, and are mistrustful of what the institutions can do once they have the weapons, they establish a number of safeguards to try to make sure that this will not happen. In my opinion, the most important of these safeguards, the most refined and complex one, is law. Law is a basic tool to make sure that the institutions will do what they are supposed to do. Law's main function is, therefore, to overcome the problems of trust that exist between the people and the institutions.

2. Law as Credibility

Trust, or rather, the lack of it, is therefore the basic problem in the relationship between the people and the institutions. As Elster puts it, trust is an essential element in human intercourse, it is 'the lubricant of society'. Without trust, cooperation is impossible, and without cooperation we get back to a state of nature. We therefore need trust-giving technologies.[24]

[22]'Hereby it is manifest, that during the time men live without a common power to keep them all in awe, they are in that condition which is called war; and such a war, as is of every man, against every man'. T. Hobbes, *Leviathan* (Oxford: Oxford University Press, 1998), chapter XIII, 8.

[23] 'It therefore appears to me incontestably true, that not only governments did not begin by arbitrary power, which is but the corruption and extreme term of government, and at length brings it back to the law of the strongest, against which governments were at first the remedy, but even that, allowing they had commenced in this manner, such power being illegal in itself could never have served as a foundation to the rights of society, nor of course to the inequality of institution'. J.J. Rousseau, *A discourse upon the origin and foundation of the inequality among mankind* (electronic resource) (Farmington Hills, Michigan: Thomson Gale, 2005), 30.

[24] See: A. Estella de Noriega, 'Law as Credibility: an outline' *Il Foro Napoletano*, I, 40-63 (2015); Id, 'Law as credibility: the case of Presidential Term Limits' 2(1) *European Journal of Legal Studies*, 116-142 (2008).

From this perspective, systems such as morals, religion, customs, politics, and even culture acquire a different light. They can be seen[25] as different technologies that humans have created to overcome, or at least mitigate, the problems of mistrust. Law is therefore, as has been said before, one of these technologies, and in fact, it can be seen as the most sophisticated technology that man has ever created to deal with trust problems. The point is that when law performs this role well, it has no paragon with its alternatives.

The game between the institutions and the people therefore starts with a promise that is made by an institution. To simplify things, let us assume that this 'institution' is the President of a State, Abraham Lincoln. As is known, Lincoln once said that 'bad promises are better broken than kept'. In my opinion, this statement exemplifies very well the origin of the mistrust of the people towards the institutions. It may be more pragmatic (and even just) that bad promises be broken rather than kept but, then, a question immediately arises: who decides when a promise is a bad promise? The answer is clear: for Lincoln, it is the one who made the promise – himself.[26] Precisely because the people know this – that political institutions can use this subterfuge to break promises – they mistrust them. Therefore, in Time 1, political institutions use law to convince the people that their promises are good promises, and that they will stick to their commitments. If in Time 2 they repent, too bad: they made a law, which is a way to objectify whether the promise was actually bad or good. Now, with their commitment encapsulated in a norm, it is no longer for the political institution to decide whether that promise was bad or good. It is for the legal system to decide it, and in the last instance, to its judicial bodies, who are, in turn, the servants of the people.

Let us develop to some extent several points that have been made before. The first thing to be said is that for this structure to work, law has to have technical superiority as a trust-giving technology if compared to its alternatives (morals, religion, etc). I give a number of examples to make clear how this model works. For example, if I am an anti-abortion politician and I promise that if I win the next election, I will do my best to ban abortion. To make my promise credible, I publicly show my Catholic faith and I let the people see me going to church every Sunday. As leader of the party, I could have included a point in my party's platform saying

[25] I owe this point to Philip Pettit. They 'can be seen' as institutional technologies, which of course does not bar other interpretations and constructions of what they are.

[26] As it clearly stems from Lincoln's own words: 'As to sustaining it, my promise is out, as before stated. But, as bad promises are better broken than kept, I shall treat this as a bad promise, and break it, whenever I shall be convinced that keeping it is adverse to the public interest. But I have not yet been so convinced'. For referencing, see n 1 above.

that I will ban abortion once in power. However, I feel I do not need to do so: people will trust me due to the fact that they all know about my Catholic fervour. In this setting, morals, or religion, play a wider role than politics in making commitments more credible.

A second example, in which law is involved, would be the example of presidential term limits. As is known, in the US, presidential term limits are established by the Constitution. However, in Spain, they are not. In the US, once a President finishes her second mandate, there is no discussion about whether she will run for a third term: the US Constitution prohibits it. However, in Spain, presidential term limits are a political commitment which is, with time, becoming a sort of political custom. However, when the second term of a Primer Minister is coming to an end, a debate arises as to whether she will run for a third mandate. This was the case with Aznar (who established the term-limitation precedent) and with Zapatero (who followed it). In fact, until a successor is appointed, the debate goes on. What this example shows is that law is a better alternative than political pacts, or political customs, in giving credibility to term-mandates limits. The bottom line is that if law is involved as an instrument to give credibility to commitments, and another instrument is preferred, then this would be a signal that law is not performing its function well. In other words, this would be a signal of a legal system malfunction.

A second point that is important to be made in this context is that, contrary to what is usually assumed by political scientists who have considered the topic of law as credibility,[27] credibility through law cannot be made equivalent to rigidity. True, law is preferred to its alternatives (morals, religion, politics, and so on) because it is more rigid than them. But inside the legal realm, the rigidity element is not enough to explain why there are different kinds of norms, and also a normative hierarchy among them. If law were equal to rigidity, without any qualification, then it would only make sense to have one kind of norm – for example, constitutional norms. However, as we lawyers know, there is a rich normative variety in developed and modern legal systems. From this perspective, the question is: if a political actor were to encapsulate a given commitment in a norm, why should she choose the constitution, or a law, or a regulation for this purpose?

The answer to this question is that credibility does not stem only from rigidity, but, more nuancedly, from an equilibrium between rigidity and flexibility. An equilibrium is defined as a situation in which there are no

[27] See, for example, R.D. Keleman and T. Teo, 'Law and the Eurozone Crisis', APSA 2012 Annual Meeting Paper (2012), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2107426 (last visited 24 May 2016).

incentives to change.[28] Further, the fact that different kinds of norms establish different equilibrium points between rigidity (R) and flexibility (F) is what explains the very existence of – and the need for – different types of norms in the legal system.

From this perspective, the Constitution, which is at the top of the normative pyramid, is the most rigid norm and the least flexible one; laws would be less rigid than Constitutions and more flexible, and so on and so forth down the normative pyramid. An important insight which is derived from this model is that the least rigid and most flexible form of law would be private law (contracts). However, to be law, all norms have to fulfil the following condition:

$$R > F$$

This means that even the most flexible form of norms, which are private law norms (contracts), still have to be more rigid than flexible; otherwise they would not be norms. If they were more flexible than rigid, we would be in the realm of morals, politics, religion and the like. I shall leave for another moment the crucial question of what the relationships are in terms of rigidity and flexibility of these other alternatives to law, but it is obvious that an answer to this acute problem cannot be given in general terms: it all depends, among other things, on the context, and above all on the cultural context, in which morals, politics, religion and the like take place.

V. Counterfactual Legal Interpretation

[28] I am thinking here of a Pareto equilibrium. Pareto optimality is a measure of efficiency. In a game composed of a determined number of players, an outcome is a Pareto optimum if there is no other outcome that would make all of the players at least as well off as in the previous outcome, and one better off than in the previous outcome. Therefore, collectively, there are no incentives to look for another outcome. If a situation is a Pareto equilibrium, then it can only be modified by hurting at least one player. It follows that a Pareto-improvement denotes a situation in which at least one player can be made better off with the new outcome (while the rest of the players remain the same). It is important to differentiate a Pareto equilibrium from a Nash equilibrium. A Nash equilibrium is an outcome in which all the players in a game have no incentive to modify their strategy, if taking into account the strategies of the rest of the players in the game. Therefore, an outcome can be a Nash equilibrium without being a Pareto optimum. Many times, Nash equilibriums are sup-optimum equilibriums from a Pareto efficiency perspective. This means that although there is room for improvement collectively, there is no room for doing it individually, if taking into account the available strategies of the rest of the players. For a very good and general summary of both concepts see K. Binmore, *Game Theory. A very Short Introduction* (Oxford: Oxford University Press, 2007).

1. Theoretical Aspects

To summarise the main points that were made in the previous section, we may conclude by saying that in a rationality setting, the main purpose that law serves is to give credibility to the commitments that political institutions make. Further, credibility can be defined as the equilibrium point between rigidity and flexibility that is found in any given norm. If the norm is an equilibrium from this perspective, it will be credible. If not, it will not be credible and therefore it will be subject to challenge.

This is the point upon which our discussion on counterfactuals impacts. In this setting, in a law as credibility setting, the main function that judges play is to try to ascertain if a given norm is an equilibrium or not. Remember that an equilibrium was defined before as a situation in which there are no incentives to change. To do this judges use a number of techniques of legal interpretation. The point is that one of the techniques is, or should be, counterfactual legal interpretation, but understood in a broader sense than what was shown in Section 3 above. This means that counterfactual interpretation should be done not only in the realm of evidence-finding and -giving, but, more broadly, in the realm of the factual basis upon which a norm rests.

The basic structure of a norm is a factual antecedent and a legal consequent. Given a factual assumption, then the norm extracts a given legal consequent. *If F, then C*. This is very similar to the basic causal structure that we set up in Section 2 of this paper. In the legal realm, the production of a given factual assumption unchains a given legal consequent. In this sense, there is a relationship between cause and effect that is not dissimilar to the relationship between independent and dependent variables that occurs in the rest of the social sciences.

When establishing a given factual assumption, the legislator may or may not have thought of the alternative courses of action to the one chosen, in alternative factual bases. We know that it is impossible that the legislator considered all possible alternative courses of action to the one it actually chose as a factual basis for drafting a norm. But it is also clear that the legislator might have thought of many of them. This being the case, my proposal is that the primary role that the judge should play should be to ascertain whether the legislator took into account these other alternative courses of action to find out whether the norm is an equilibrium or not.

This enlarges and restricts, at the same time, the role of judges in the power game between themselves and the legislator and the executive. Based upon a certain vision of democracy, which connects with what is called neo-republicanism, government is 'of the people, by the people, and

for the people', as Lincoln said in his Gettysburg Address.[29] This means that in a representative democracy, there are many doubts as to the democratic legitimacy of judges, who are, particularly in the continental legal tradition, but civil servants who are not democratically elected. There is and there always should be a suspicion as to the role of servants of the people that judges play. Like other institutions, they have their own particular preferences, motives and strategies. But unlike other institutions, the people cannot expel judges from power as readily as they can do with the government and the parliament, which is through vote. Still, we also know that for the system to work, and in particular, for the law as credibility system to work, we need an independent institution (independent from the legislator and the executive) to make autonomous judgements on the law. This is why judges' role, when judging the laws, should be very entrenched; on the other hand, the system should also have enough flexibility for it to be stretched when it is clear that the law is not performing well its function of credibility-giver.

Judges should therefore restrict their analyses on law to whether a certain norm is an equilibrium or not. To do so, they have to inquire into the factual basis of a given norm. And one of the ways to do this – a privileged one – is through counterfactual analysis.

More in particular, the steps that a judge has to take in order to ascertain whether a given norm is an equilibrium or not are the following (see Table 1 hereinafter). Step number one would be to find out if the legislator took into account a counterfactual situation. If the answer to this question is affirmative, then the second step should be to ascertain whether the norm is an equilibrium or not from a credibility perspective. If it is, then the norm would be fit from a credibility perspective – and therefore the case would end there. If instead it is not, then the judge can indicate the solution that, according to her, would generate an equilibrium point; however, the judge can simply *indicate* this solution to the legislator, but she cannot impose it. In this case, the judge would have to show deference to the legislator and could not, for example, in the case of constitutional review of legislation, strike down the norm. It would be for the legislator to do the job and put the norm back in equilibrium.

Another possibility is that the judge finds out that the legislator did not take into account the counterfactual situation. She would then, as a second step, have to determine whether, *taking into account the counterfactual situation*, the norm is still an equilibrium. It might be the

[29] See P. Pettit, *On the People's Terms: A Republican Theory and Model of Democracy* (New York: Cambridge University Press, 2012) and Id, *Republicanism: A Theory of Freedom and Government* (Oxford-New York: Oxford University Press, 1997).

case that if the legislator did not take into account the counterfactual, then the norm is still an equilibrium; however, with a high degree of probability, this will not be the case. In other words, the fact that the legislator did not take into account the counterfactual would be a primary indication that the norm is not an equilibrium. If, in fact, the judge finds out that the norm is not an equilibrium, then, in this case, she should give the solution to the case at hand that she finds is an equilibrium from a credibility perspective. If we are in the domain of constitutional review, she should not only strike down the norm in question, but give a solution for the case which has a regulatory bearing. This is to say that the norm cannot be an equilibrium for that particular case *and only for that case*; it has to try to fix an equilibrium point from a general perspective. In credibility terms, it makes no sense to speak of such a thing as the '*giustizia del caso concreto*'. This puts another brake upon the judge's discretionality when making legislation.

2. An Example: Case Re E (A Minor) (Wardship: Medical Treatment) of 21 September 1990

Let us shed some light on the previous discussion. I will illustrate it with a particular case which arose in the English legal jurisdiction: Re E (A Minor) (Wardship: Medical Treatment) of 21 September 1990.[30] As evidence of the closeness that exists between legal counterfactuals and fictional counterfactuals, it is interesting to note that this case inspired one of the most well-known novels of Ian McEwan, titled *The Children Act*.[31] The facts of the case are simple. 'A', a minor of almost sixteen years old, was hospitalized since he was discovered to suffer from leukaemia. The doctors prescribed him a blood transfusion, which he refused due to his being a Jehovah's Witness. Because his case was desperate, the hospital applied for a Court's order to oblige him to receive the blood transfusion. The lawyer of the parents of A cited in support of the child's refusal Art 8 of the Family Law Reform Act 1969, which said:

> 'The consent of a minor who has attained the age of sixteen years to any surgical, medical or dental treatment which, in the absence of consent, would constitute a trespass to his person, shall be as effective as it would be if he were of full age; and where a minor has by virtue of this section given an effective consent to any treatment it shall not be necessary to obtain any consent for it from his parent or guardian'.

[30] Family Court Reports/1992/Volume 2/Re E (A Minor) (Wardship: Medical Treatment) – [1992] 2 FCR 219.

[31] I. McEwan, *The Children Act* (London: Jonathan Cape, 2014).

The parents' lawyer argued that since the minor was about to turn sixteen, then this article should apply to the case at hand. In fact, the minor refused to receive the treatment. Lord Ward, the Court of Appeal's judge of the case, in fact interviewed the child and he confirmed his refusal. In addition, Lord Ward did not deny that, as the minor was so close to the age of sixteen, Art 8 of the Family Law Reform Act 1969 should apply. However, in his judgement, Lord Ward said that the well-being of the minor overrode any other concern, in particular, the concern for the protection of his human dignity. More in particular, his precise words were the following:

> 'The circumstances required the court to make a decision. In doing so, the court must regard the child's welfare as the first and paramount consideration. The judgment of what the child's welfare dictated was to be exercised objectively. But regard had to be paid to the position of this child as a boy of growing maturity living in the religious society that he did. Whether or not he was of sufficient understanding to give or withhold consent was not the issue before the court. The child's wishes were an important factor and, as he was nearly 16, a very important matter which weighed heavily in the scales. He appreciated the consequences if not the process of his decision to refuse a medical intervention which might save his life. He was close to the time when he might be able to take such a decision. Therefore, the court should be slow to interfere. But it should consider to what extent his refusal of treatment was the product of his full and free informed thought. The influence of the teachings of the Jehovah's Witnesses was strong and powerful. The child might assert his views but his volition had been conditioned by very powerful expressions of faith. It could not be said that at the age of 15 his will was fully free. The risk of serious infection from a blood transfusion was infinitesimal and not a risk which would have stood in the father's way but for his religious convictions. Therefore, when balancing the wishes of the father and son against the need for the chance to live a precious life, it had to be concluded that their decision was inimical to the child's well-being. As a result, the hospital would be given leave to treat the child as they considered necessary, including the administration of blood transfusions'.

Stemming from the facts of the case, it is clear that Art 8 of the Family Law Reform Act 1969 had not foreseen this kind of situation when it was drafted. The norm was passed thinking in a very particular factual situation: that of contradictory views between the minor and her parents or guardians as regards her consent for medical treatment. The aim of the norm is therefore clear: it is to protect the will of the minor. Therefore, given

a contradiction (factual basis of the norm), more weight should be given to the will of the minor than to the decision of her parents or guardians (legal consequent).

But, what if the will of the minor put at risk her own health? And what if there was no contradiction and the will of the parents, guardians, and minor matched? This was not clearly in the mind of the legislator when the norm was drafted. In other terms, the legislator was not thinking of the Jehovah's Witnesses' refusal of blood transfusions and similar cases when it was enacted. Though Lord Ward does not say it openly, he is in fact making a counterfactual legal analysis in his decision. If this were the case, that the minor put at risk her own life in a particular case, then 'the child's welfare (would become) the first and paramount consideration'. By doing so, Lord Ward would be overriding Art 8 of the Family Law Reform Act 1969 and applying another principle, that of the child's welfare. In fact, this case, and similar cases, prompted a reform by the legislature of Section 1 of the Children Act of 1989, to take into account the principle according to which 'the child's welfare is the court's paramount consideration' in courts' decisions on children's rights.

Let us analyse this case more closely and from the perspective of game theory. In reality what we have here are two contradictory interests that cannot be weighed: that of the minor and that of the medical doctor. In the game between the minor and the doctor, the minor wants her will to be preserved, and the doctor wants life to be preserved. The judge is the third player in this game: she has to decide whether the minor's preference should be preserved or whether the doctor's preference should be preserved. As said above, the judge also has her own preferences, strategies, and motivations. She is inspired, in particular, by self-interest, which was defined before as an increment of her feeling of internal well-being. If the judge let the minor die, then she would feel bad with more probability than if the minor were to feel well. Knowing this, the judge has an interest in changing the law. Protecting the minor's will at any price is not an equilibrium. Therefore, the judge changes the law so that a new equilibrium emerges: the will of the minor shall be protected unless the minor's life is put at risk. In this game, it is very clear that at least two players (the doctor and the judge) have incentives to modify the previous legal equilibrium. The fact that the judge is supported by one of the players in the game makes it possible to make an alliance and defeat the minor. In the new equilibrium, only one player would have incentives to change, and two would have not. This means that the new equilibrium would be sub-optimum and therefore unstable: the minor could always try

to convince the legislator so that a new, bolder norm, protecting minors' will, be passed.[32]

From a more general perspective, we could say that the problem with the previous legislation (with Art 8 of the Family Law Reform Act 1969) was that it was not credible enough. Had the legislator made the counterfactual, integrating a situation in which the life of a minor was put at risk by her own decision, then it is clear that a judge would have tried to protect her life even 'from herself', as Ward puts it in his judgment. In other words, it is not very credible to think that a judge would have simply relied on the minor's will to adopt a decision that would surely have endangered her own life.

VI. Conclusions

I will not attempt to summarise what has been argued above. The main argument can be, instead, simplified in the following way: counterfactual thought experiments should not only be limited to the evidentiary phase of a trial, but should be enlarged to try to ascertain whether a norm is or is not an equilibrium from a credibility perspective. Finally, it would be important not to see counterfactual thought experiments in law as a blank check for judges to substitute the judges' preferences for the legislator's discretion. If correctly understood, counterfactual thought experiments can enlarge judges' margin of manoeuvre as much as they can restrain it, as the final outcome of the case analysed in Section 5, above, seems to suggest.

[32] The minor would have a second alternative, which would be to wait (if he could) until he is not a minor anymore, when he turns eighteen. Thus, according to the UK legislation, in this case the situation is very clear and the will of the adult has to be preserved, even if it puts at risk her own life. In fact, and in an unprecedented dramatic turn, this is what happened in the case at hand: 'A' suffered from a new illness a time after, and he refused, again, a transfusion. He died, accordingly. This proves that in this case, the new equilibrium was sub-optimum. The full story is told here: http://www.theguardian.com/books/2014/sep/05/ian-mcewan-law-versus-religious-belief (last visited 24 May 2016)

TABLE 1: COUNTERFACTUAL THOUGHT EXPERIMENTS AND LAW