

# Ought Computers Adjudicate?

Aernout Schmidt\*

## 1 Introduction

In 1991 Jaap van den Herik published his seminal inaugural address as a professor of legal informatics at Leiden University.<sup>1</sup> He discussed the question whether computers can adjudicate. Not Van den Herik's arguments were feverishly contested at the time, but his audacity to bring the question onto the Dutch legal academic fore at all. Ever since, it has been his adopted mission to emphasize over and over again that computers are and will be capable of much more than we are willing to imagine. He is right. And he is also right in expecting our societies to be caught unaware if we do not take this phenomenon as a serious matter of concern.<sup>2</sup>

Van den Herik's inaugural lecture ended by a double note. One note gained the larger part of the public's attention and concerned the proposition that computers *will be capable* of adjudicating in a better way than human judges will do in the future. This may be a question of several centuries (he tentatively mentions the year 2984), but it will happen eventually. Van den Herik is willing to take any bet on it. And as he has already won his bet with Hans Böhm on a computer gaining victory in a serious game of chess against the reigning world champion, few of us dare take him on.

As predicting the future may be interesting entertainment, it is not the tack I choose to pursue. Here, I want to take up the other note by Van den Herik as stated in 1991. If we do accept, as a working hypothesis, that computers are capable of adjudication, then we have to address the question whether we let them do so, or even promote them to adjudicate. In brief, assuming that computers are capable to adjudicate, we face the new question: ought they? Ought computers adjudicate if they can? This is the question I want to address in my essay. And by doing so, I would like to pay tribute to the man who set me on this track.

---

\* Aernout Schmidt is a full professor of law and IT at the Department of Metajuridica of Leiden University. Email: [a.h.j.schmidt@law.leidenuniv.nl](mailto:a.h.j.schmidt@law.leidenuniv.nl).

<sup>1</sup> H.J. van den Herik, *Kunnen Computers rechtspreken?* Arnhem: Gouda Quint bv 1991.

<sup>2</sup> As it is always dangerous to phrase any thought, I want to stress that I am trying to avoid technical terminology. Consequently, any resemblance of my words with technical terms is coincidental. "Matters of concern," for instance, is meant to mean what it means to people who manage without reading Bruno Latour.

## 2 Conjecture

Let me hasten to state that the foregoing does not imply that I agree with everything in Van den Herik's inaugural address. One of the reasons to address the current question is the wording Van den Herik chose for his concluding proposition: "after three months of impeccable computer-adjudication advice, the computer becomes the adjudicator, irrespective of what each one of them may think about the matter" (p. 33, my translation). This is not a speculation about what computers are capable of, but about how complex social institutions will react to challenges posed by innovation. Admittedly, I am not a neoclassical economist, but I doubt that the implicit reasoning — which seems to be related to the efficiency-effectiveness brand — is valid. Looking at the chess scene, for instance, I do not see any analogous prediction to become true. Although computers are capable of high-level chess-playing and although computer-chess competitions do exist, there is little evidence that "people-chess" competitions are being displaced by computer-chess competitions at any level or scale. Not even at those levels where generally affordable chess programs are available. So, I venture to suggest that there might be other than economic reasons that prevent our natural laziness to take over and let computers run chess competitions on their own, on our behalf, thus setting us free to do more important things. Apparently, some values are simply not fit for delegation. Individual, existential pleasure (as may be involved in playing a game of chess) may be one of them.

The *topos of ignorance* I want to address encompasses many interesting questions of which I would like to single out the following: are there inherent aspects in our adjudicatory arrangements that will prevent us from accepting agency (proxy, delegation) relationships with computers to act as judiciary on our behalfs? My working plan is as follows: (1) specify the domain of discourse and (2) discuss several imaginary "refutations" against the proposition that computers ought to adjudicate as soon as they consistently outperform human judges. We have, after all, no better scientific approach available than to frame our ignorance in concepts, look around us, concoct "risky" (falsifiable) conjectures, and apply adequate observations, valuations, and reasoning methods, in search of refutations.<sup>3</sup>

---

<sup>3</sup> K. Popper. *Conjectures and Refutations*. London: Kegan Paul, 1963. Although almost anywhere else accepted as a sound (empirical) scientific approach, in legal academics it is quite eccentric — or rather, it has somehow become almost entirely obsolete nowadays. This trend is very difficult to understand when we take into consideration that a substantive part of legal scientific and philosophical work still in print and read seriously rests on empirical observation and argument (e.g., Hobbes (*Leviathan*, 1652), Montesquieu (*De l'Esprit des Lois*, 1748), Rousseau (*Du Contrat Social*, 1762), Smith (*An Inquiry into the Nature and Causes of the Wealth of Nations*, 1776), Bentham (*An Introduction to the Principles of Morals and Legislation*, 1789), Mill (*On Liberty*, 1859), Popper (*The Open Society and Its Enemies*, 1943), Hart (*The Concept of Law*, 1961), Fuller (*The Morality of Law*, 1963) and Rawls (*A Theory of Justice*, 1971)). Perhaps as a result of the current dominance (and among legal scientists common misreading) of the "naturalistic fallacy" that is generally

### 3 Ought Judges Adjudicate?

It is no more than reasonable to start and approach the question “*ought computers adjudicate?*” in a traditional conceptual manner, which immediately suggests that we expect to find applicable conceptualizations in traditional arguments. This brings our attention to a question so seldom asked, that it may produce unpredictable reactions in the reader’s mind. The mere formulation of the question “*ought judges adjudicate?*” may at first sight be considered irreverent, outlandish or downright dangerous — but never irrelevant in this context. If we know why judges ought to adjudicate, we also know some relevant functional requirements that may help to decide the issue under which conditions we ought to let computers take over.

So why is it, that in the legal discipline (of which I am now for more than 35 year a member) I have never heard this question at all, and, consequently, never heard it seriously discussed? Almost all professional legal attention is directed towards issues of *how* judges should adjudicate, not whether *judges* should do so. The proposition that judges should adjudicate is one of the fundamental assumptions the legal discipline is built upon. The very question shakes this assumption and answering it may consequently shake the entire discipline. For a discipline averse to being shaken, the question is not attractive. And whether judges should adjudicate is a dangerous question indeed. Equally dangerous is the question whether computers ought to adjudicate; perhaps the latter is even more dangerous, as we might reach in theory the conclusion that neither judges nor computers are fit for adequate adjudication. Certainly you might agree — I assume — that it would not only create the risk of some serious additional instability in our legal arrangements, it would also raise the related question: why do we need adjudicatory functionality to be organized in our societies at all? Let me start from there.

### 4 Framing Adjudicatory Organization

Interestingly, the most general argument on why we need organization at all can be harvested from an economic publication by Ronald Coase (1937)<sup>4</sup> — from an article so beautifully simple that it took more than 50 years before it drew sufficient attention to assign him a Nobel Memorial Award — in 1991 as a matter of fact, the very year of Van den Herik’s inaugural lecture. The argument is almost self-evident once the right question is available. Coase’s question concerned the

---

considered to flaw any attempt to derive that what ought to be from that what is, and rightly so (Hume/Moore). What the legal discipline seems to forget, however, is that this naturalistic fallacy is equally valid for any attempt to derive anything that will be from what is (or has been) — and that Popper’s approach to the growth of science, which is surely an empiricist one, was designed to avoid it. For who, as I do, accepts the assumption that what ought to be may be formulated as a (legal) behavioral theory, Popper’s approach is equally valid.

<sup>4</sup> Ronald H. Coase, The Nature of the Firm, *Economica*, 4 (November 1937): 386-405.

existence of hierarchically organized firms in the face of the fundamental economic assumption that only the market (the space for unorganized exchange) is optimally efficient. How can it be that we observe some organizations to be more efficient than the unorganized market, if we assume the market to provide optimal efficiency? Coase suggests that the costs of knowledge gathering and consolidation, of specialization, of agency, and of long-term agreements may be less than those involved in ‘spot market’ transactions, where each and every transaction is burdened by the need to fulfil cumbersome information requirements and to (re)do comparisons and evaluations. As a result, organization has begun to matter in economic models<sup>5</sup> and the assumption that theories of market efficiency may ignore information costs (and may naively accept the two-part assumption that each market player (i) is omniscient of the market’s supplies and (ii) has a peculiar form of amnesia concerning earlier dealings) has been shaken considerably.

So far, so good. Coase provides an argument for legitimacy of *economic* organization. Is this argument also fit for a fruitful discussion on the legitimacy of *judiciary* organization? Can the argument be generalized?

Coase’s argument may very well be generalized as follows. Organization may be preferred to individual *ad hoc* behavior, when the overall value-return provided by the organization outperforms the overall value-investment, required to organize. As soon as this is no longer the case, an organization loses its overall legitimacy and its existential foundation. This generalization suggests that there may be different reciprocity “equations” for different value types that are linked to organizational goals. “Not just for the money” as Bruno Frey suggests<sup>6</sup> and as the persistent existence of lower-level “people-chess” competitions shows. In the approach suggested, the values exchanged in the overall reciprocity equation of a chess competition would be of the form (1) against (2), viz. (1) the aggregated values of “the opportunity to play regularly against interesting players” against (2) the aggregated values of “individual fees.” This equation would not be influenced by “computer-chess” competitions at all as these simply do not occur in the valuation equation concerned.

The complexity and the vagueness of my statements above make me wonder whether the readers are still with me. Therefore, I will try to rephrase my statements in the form of a semi-formal model. My discussion aims at an appropriate conceptualization for scientific refutation attempts against the proposition that computers ought to adjudicate if they perform consistently better than human judges. I argued earlier that this should be a question about agency (proxy, delegation) and hierarchical organization. I then suggested that there are value costs and value benefits in a reciprocity relation between any organization (acting on a public’s behalf) and that public. I showed that these values need not necessarily be fit for proxy and that they need not always be expressible in numerical variables. Moreover, I suggested that reciprocity relationships express

<sup>5</sup> E.g., O. E. Williamson, Comparative Economic Organization: The Analysis of Discrete Structural Alternatives, *Administrative Science Quarterly*, 36, 1991, p. 269-296.

<sup>6</sup> Bruno S. Frey, *Not Just For the Money*, Cheltenham: Keagan Paul, 1997.

the legitimacies of organizations, thus potentially offering a decision procedure (if adequately instantiated) to decide any question regarding a proxy. Let me try to outline the decision procedure and summarize the foregoing in a model. I will use four equations, one for overall or aggregated legitimacy of an organization and three for internal or local legitimacies.

The overall, aggregated legitimacy  $AL_{OC}$  of organization  $O$  in community  $C$  may be expressed in an equation, relating (1) the aggregated organizational value return  $f[R]_{OC}$  and (2) the aggregated organizational value investment:  $g[I]_{OC}$ :

$$AL_{OC} = f[R]_{OC} - g[I]_{OC}, \quad (1)$$

where an organization loses overall legitimacy as soon as  $AL_{OC} < 0$ .

It should be stressed that the aggregation functions  $f$  and  $g$  are specific to the OC-contingent variable sets  $[R]_{OC}$  and  $[I]_{OC}$  (which represent the “data models” framing the values that describe the value returns and the value investments). These values need not be expressible in dollars. [Of course this raises a problem. It will be discussed later.] As I am now at the stage of formulating some semi-formal conceptualizations for organizational legitimacy, I want to address how the other existential aspect of legitimacy may be modeled. I refer to the concept of internal or local legitimacy. Local legitimacies (LL) concern equations that venture to express the related notions of organizational fairness, distributive justice, and corruption. It recognizes that even positive overall legitimacy may internally show such inequality in the distribution of investments and returns over the executives (E), the other functionaries (F), and the publics (P) of an organization, that it may fall apart due to loss of internal legitimacy. For this argument I accept three local legitimacies, one for the executives of the organization:

$$LL_{OCE} = f[R_E]_{OC} - g[I_E]_{OC}, \quad (2)$$

one for the functionaries of the organization:

$$LL_{OCF} = f[R_F]_{OC} - g[I_F]_{OC}, \quad (3)$$

and one for the public of the organization:

$$LL_{OCP} = f[R_P]_{OC} - g[I_P]_{OC}. \quad (4)$$

I complete the decision procedure by adding the requirement that an organization loses local legitimacy if one of the equations (2) to (4) has a negative value. The reason is self-evident. Any hierarchical organization needs a management, needs a work force, and needs publics (customers, consumers, members, citizens, subjects, believers, whatever). So, any hierarchical organization will lose local legitimacy (or part of its overall legitimacy) when any one of them loses legitimacy. I consider any argument to be a refutation when it shows that any of these equations (concerning judiciary organization in a jurisdiction) necessarily has a negative outcome when instantiating computers for judges.

## 5 Refutations

Let me now proceed by discussing some imaginary refutations. I have selected four of them. The first two (*Against the law? Against the Rule of Law?*) follow the lines of mainstream legal scientific practice and claim to lead to the conclusion that adjudication by computers ought not be allowed as it is against both the law and the rule of law. The third one (*Inherent incommensurability?*) claims that the values in our reciprocity equations are inherently incommensurable and thus prevent adequate algorithmic modeling. The last one (*An unavoidable trap?*) claims that judges generate propositions about what ought to be, and that replacing them with model-based computing will inescapably involve application of the naturalistic fallacy.

### 5.1 Against the Law?

Mainstream legal argumentation will look at valid law for answers to ought-to questions employing valid legal reasoning. As it is impossible to find valid law allowing computers to adjudicate (as, for instance, Dutch law implicitly requires judges to be human by explicitly requiring an affidavit of decent behavior before anyone can be appointed as a judge), we seem to encounter a serious refutation. However.<sup>7</sup>

The legal reasoning involved needs to be valid and what lines of legal reasoning are valid has been researched and commented upon extensively. It is not my aim to enter into the quagmire of legal-academic dispute in this area of jurisprudence. For making my point here I need some practical characteristics to break down the field heuristically, so that I can make use of a generalized legal approach. I use two dimensions for this, irreverently and mercilessly<sup>8</sup> borrowed from the work of the constitutional law scholar Bobbitt<sup>9</sup> and the psychologist Fodor<sup>10</sup> respectively. Bobbitt analyzed the types of arguments used during the history of constitutional review in the USA and comes up with a limitative<sup>11</sup> list: historical (addressing the intentions of the rule makers), textual (addressing the present sense of the rule wording), doctrinal (addressing the principles and relations derived from precedent), prudential (self-consciously addressing the role and position of the party or institution that is ruling), structural (addressing the implicit meaning from the existence and role of legal institutions) and ethical (arguments addressing the *ethos* of a<sup>12</sup> jurisdiction or *politeia*).

<sup>7</sup> The following part (in smaller typeset) is an auto-citation from a contribution (together with H. Franken) to H. Snijders and S. Weatherill (eds.) *E-commerce Law*, Kluwer Law International, 2003 (p. 123-124).

<sup>8</sup> Irreverently, because put to use in a very blunt, practical and generalizing way that does not and cannot respect the finesses of their fine work.

<sup>9</sup> Philip Bobbitt, *Constitutional Fate — theory of the constitution*, Oxford University Press 1982.

<sup>10</sup> J.A. Fodor, *Psychosemantics*, MIT press 1987.

<sup>11</sup> Despite his claim not to be exhaustive on p. 8, the exercise proposed on pp. 94 and 95 suggests the opposite (Bobbitt, *o.c.*)

<sup>12</sup> Here I generalize Bobbitt's typology — he mentions the *ethos* of *American* polity.

Fodor is interested in natural and in artificial intelligence and in how human knowledge representation relates to knowledge representation in IT. He finds the distinctions of abstraction layers as a basic similarity and considers the positioning of abstraction layers relative to reasoning of fundamental importance. Imposing Fodor's abstraction layers upon legal argument I suggest six layers: common beliefs (like in a free market, the ten commandments, being responsible for one's acts), principles (like in human rights, in the constitution), rules (like in legislation, for instance democratically formulating exceptions to human rights), policies (like in administrative plans, interchange agreements and common-law precedent), individual judgments (like in civil-law precedent, in transactions) and facts (making up the facts of a case).

Types of argument can be related to levels of abstraction in the framework as presented in Table 1. The framework may help to analyze legal reasoning and to present a heuristic distinction between valid and invalid legal reasoning in the different contexts. First of all, the different levels of abstraction represent different sources for legal arguments. In mainstream legal analysis, only facts, judgments, policies, rules and principles are valid sources for argument, and only that legal reasoning is valid that brings facts within the constraints of valid judgments, that are within the constraints of valid policies, that are within the constraints of valid rules, that are within the constraints of valid principles. In normal circumstances, common beliefs do not provide valid sources for legal argument (indicated in Table 1 by  $\diamond$ ). Moreover, ethical arguments are also not admissible (indicated by  $\diamond$ ) — anyway, within a stable *politeia*, arguments at the common-belief level and arguments of the ethical type are most often coherent with valid legal reasoning.

**Table 1.** A framework for analyzing legal reasoning

	History	Text	Doctrine	Prudence	Structure	Ethics
Beliefs	$\diamond$	$\diamond$	$\diamond$	$\diamond$	$\diamond$	$\diamond$
Principles						$\diamond$
Rules						$\diamond$
Policies						$\diamond$
Judgments						$\diamond$
Facts						$\diamond$

Now the main flaw in the refutation suggested concerns the assumption that the legal rules and policies of today will remain valid in the future. Legal rules and policies are adapted all the time, however. The argument that today's law will rule tomorrow's world is simply inaccurate. And, in the reciprocity equation where the value of organized scientific argument is weighed against the investment to read, argue and be convinced, inaccurate arguments are not sufficient.

## 5.2 Against the Rule of Law?

The second refutation expresses the concern that computers do neither have personality nor morality, and that we consequently need human judges to guard the morality of our legal systems. This refutation stems from the second perspective in legal academics: jurisprudence or legal theory. It is of the meta-juridical type and focuses on the quality of legal systems as such. Its results<sup>13</sup> yield (the discussion of) existential conditions for legal-system quality and may be seen as recommendations which provide specifics for the reciprocity equations mentioned earlier.

The Rule of Law and its moralities are central issues. A practical approach towards an operationalization of legal-system morality can be realized by instantiating the *rule of law* with Fuller's list<sup>14</sup>, supplemented by two additional requirements, on system-role organization<sup>15</sup> and on reciprocity.<sup>16</sup> Let me explain somewhat further:<sup>17</sup>

The Rule of Law refers to a working, minimumly decent legal system in a minimumly decent society. As Tamanaha shows in his recent book, the 'Rule of Law' concept is currently not only a very popular one, it is also widely abused as an argument by governments of various reputation to expect civilians to comply with their rules.<sup>18</sup> On top of that, it is an elusive concept. Almost everyone using it means something different. I need a working notion and thus start with the question of which normative systems are minimally decent law systems. Fuller (1969) specifies eight necessary moralities of duty in this context:

- (a) there cannot be a minimally decent law system if it does not have any general rules at all;
- (b) there cannot be a minimally decent law system if the rules are unavailable to the addressee;
- (c) there cannot be a minimally decent law system if legal practice shows abuse of retroactive legislation;
- (d) there cannot be a minimally decent law system if the rules are not understandable to the addressee;

<sup>13</sup> E.g., Lon L. Fuller, *The Morality of Law*, revised edition, Yale University Press 1967, John Rawls, *The Law of Peoples*, Harvard University Press 1999 and Brian Z. Tamanaha, *On the Rule of Law*, Cambridge University Press 2004.

<sup>14</sup> See Lon Fuller, *The Morality of Law*, (1967 ed.) and Brian Tamanaha, *On the Rule of Law*, (2004).

<sup>15</sup> The requirement of adequate system-role organization (separation of powers à la Montesquieu) is, after all, a cornerstone of Western jurisprudence. It is consequently an important concept in (formal) comparative analysis.

<sup>16</sup> Reciprocity is considered fundamental in virtually every (Western) political and socio-economical theory, including those proposed by Hobbes, Hume, Locke, Smith, Fuller, Hayek, Rawls, Coase, Searle and Pessers.

<sup>17</sup> The following part (in smaller typeset) is an auto-citation from a contribution to Aernout Schmidt, Wilfred Dolfsma and Wim Keuvelaar, *Fighting the War on File Sharing*, The Hague: T.M.C. Asser Press 2007 (p. 159-161).

<sup>18</sup> Brian Z. Tamanaha, *On the Rule of Law*, Oxford University Press, 2004.



- (e) there cannot be a minimally decent law system if the enactment of rules is contradictory;
- (f) there cannot be a minimally decent law system if rules require conduct which is beyond the powers of the addressed party;
- (g) there cannot be a minimally decent law system if rules are subject to changes so often that subjects cannot orient their actions thereto;
- (h) there cannot be a minimally decent law system if there exists incongruence between the rules and their administration.

I picked Fuller's list because it conveys a well-founded, non-controversial set of formal conditions that can be used as a simple checklist for the existence of a minimally decent law system. Fuller's list has nevertheless been seriously criticized as underspecific. I agree and add another non-controversial condition to Fuller's list: the condition of adequate institutional roles /'a la Montesquieu:

- (i) there cannot be a minimally decent law system if it does not show in its elites the three interdependent roles (legislation, adjudication, administration) as addressed in the maxim of the separation of powers.

These nine conditions have been considered to provide a formal description of the Rule of Law. Still, they have as such been subject to additional criticism, mostly because they would allow some very unsavory systems inside the domain of minimally decent law systems. Again, I agree and will consequently add one other, non-controversial condition to the list:

- (j) there cannot be a minimally decent law system if it does not show at least one effective procedure for feedback, supporting reciprocity in the relationships between the law system's elites and the law system's subjects.

What we have thus far is the definition of what I consider to be a minimally decent legal system, and as such it is an articulation on paper. I need the following closing condition:

- (k) there is a minimally decent law system if, and only if, it shows itself in accordance with conditions (a)–(j) *in practice*.

By now, I have some background to confront the refutation that since computers do not have morality, they cannot function as a legal-system elite that guards legal-system morality. The theory suggests that some form of minimal legal-system morality will be the emergent result of complying with the 10 formal rules mentioned above. Of these, only five relate to the judiciary: (c), (e), (h), (i), and (j). The first three of these can hardly play a role in the refutation under discussion as computers are considered functionally superior to humans where it concerns rule-compliant (c) or consistent behavior (e, h).

The fourth formal rule protecting minimal morality of our legal systems (i) requires a judiciary that is sensitive to legislative feedback and does not a priori stipulate anything preventing computers to play the part. The reasoning behind the separation of powers, however, does require the behavior of the judiciary to be sensitive to legislative input and to be receptive to feedback information. As a consequence, if computers are to play the role of judiciary, they must be designed in a manner allowing the legislator to change their behavior accordingly. As solutions are conceivable (e.g., designing intelligent maintenance algorithms or by appointing what Bovens has coined "system bureaucrats") I do not consider this a valid refutation. [The argument that it will be impossible to foresee what these

inputs will be (and it thus will be impossible to model adequate maintenance algorithms) is outside our domain of discourse as it is our working hypothesis that computers in fact can completely replace the human judiciary.]

The last formal rule protecting minimal morality of our legal systems (j) requires that there exists at least one feedback relation between the judiciary and the adjudicated. The related refutation often refers to the essential need for empathic relationships in adjudicatory arrangements. In this line of thought, adjudicational arrangements are inherently real-time and interactive. The refutation is interesting, but hardly persistent. In the current practice we already see that these requirements are being adapted to the emerging IT-possibilities. An even stronger argument recognizes that empathy is hardly to the point for the feedback requirement. A poorly motivated judicial opinion is poor feedback, irrespective of whether or not some empathy is shown by the judge. This 'Rule of Law' refutation does not prevail as the value of empathy is secondary to the value of feedback accuracy.

### 5.3 Inherent Incommensurability?

In order to be able to employ the decision-procedure equations, we need to know what values are parts of them. As suggested earlier, the economic or monetary perspective may be inadequate, as we may even want to delegate adjudicatory authority to human organizations without experiencing any direct financial return. I, for instance, have never in my life had the need for any court's opinion. And although I know quite well that part of the taxes I pay are spent on our judiciary, I do not feel any lack of fairness here. After all, the very existence of a decent court system has value, and may even be considered a necessary condition for public stability in a society, where corruption and crime are kept at large and where families, cultural institutions, and economic markets may flourish. Thus, I am willing to contribute not only financially to our legal system, but also in freedoms — by refraining from unlawful acts and by accepting judiciary authority. Now, if economic organizational exchange can be measured with a monetary yardstick (prices and welfare), adjudicatory organizational exchange might consider the balance between the values of authorities and freedoms transferred on the one side and the values of power-application practices and securities returned on the other side.<sup>19</sup> To let computers provide quality adjudication we need to be able to model explicitly the values included in the reciprocity equations. And we can't, can we? We do not yet know how to design an explicit yardstick that makes the values of freedoms, authorities, power-application practices, and securities commensurable. Here, economics has so great an advantage over, e.g., the legal, political, anthropological, and theological disciplines, that non-economic arguments draw little attention in a political discourse nowadays.<sup>20</sup>

This phenomenon — the intrinsic non-quantifiability of some value types — induces some wonderful reactions. Wittgenstein-I argued not to speak about

<sup>19</sup> See for a related approach Mancur Olson, *Power and Prosperity*, New York: Basic Books 2000.

<sup>20</sup> Except when focused on fundamentalist behavior.

them, as we cannot know what we cannot quantify and as we better not speak about what we do not know. Later, he retracted this position. However, the Wittgenstein-I approach has proven to be very alluring in practice. It fits, for instance, common attitudes in both econometrics and informatics. As they cannot process non-quantifiable or unknown information, they tend to deny or underestimate its value and relevance. Kaplan and Shavell's *Fairness versus Welfare*<sup>21</sup> is an interesting point in case. They claim that if fairness has value at all, it should be made part of economic modelling and that considering the value of fairness outside these models will lead to sub-optimal welfare. Their slogan might be: what we cannot quantify is bad. In my opinion this proposition has a strong family likeness to the proposition that we better refrain from speaking about things we cannot quantify or know. A more consistent approach seems to me to rephrase these propositions in a manner that has a smaller tendency to place the discipline involved in the center of the world, which would presumably be a better place when economists refrained from speaking about values they cannot handle in their theories in stead of suggesting that these values are nonexistent, not important or inefficient, and when computer-service designers would refrain from speaking about values they cannot handle in their programs in stead of suggesting that these values are nonexistent, not important or inefficient. Let me quote Bruno Frey here: "This 'new market orthodoxy' by far stretches the limits in which the price system is an effective and useful social decision-making mechanism (...) It is indeed neither possible nor desirable to build a society solely or even mainly on monetary incentives — nor, of course, on commands and regulations."<sup>22</sup>

Still, as a refutation against the proposition that computers ought to adjudicate if they can, these considerations are beside the point because they are outside our domain of discourse, as it is our working hypothesis that future research will result either (i) in specifying a form of commensurability of relevant values or (ii) in gaining some other effective approach to superior adjudication results, as part of the then available capabilities that will support computers to adjudicate in a better way than humans.

#### 5.4 The Naturalistic Fallacy: an Unavoidable Trap?

There is a rather serious philosophical and very general refutation against the proposition that computers ought to adjudicate. It concerns the so-called naturalistic fallacy (the is-ought derivation problem) and reads as follows. Scientific knowledge concerns what is. What ought to be cannot be derived from scientific knowledge. If we know enough to make computers behave in a manner that outperforms judges, they will necessarily process knowledge (propositions) into norms (prescriptions). Computers that adjudicate thus cannot escape the naturalistic-fallacy trap. Is this a serious refutation attempt? Does the argument not as easily apply to human judges, too?

<sup>21</sup> Louis Kaplan and Steven Shavell, *Fairness versus Welfare*, Harvard Law Review 2001.

<sup>22</sup> Bruno S. Frey, o.c., p. 2.

We observe actual behavior that can be understood to domesticate rationally the manipulation of mathematically incommensurable values all the time. But then we are not in the realm of science, but in real life, employing practical reasoning. As any chess-player may corroborate, people seem to succeed in making choices and decisions about values they cannot measure, concerning issues they can only know partially — even under time-constraint time constraints. They often even enjoy it. Ask any individual to come up with “grades” between 1 and 10, expressing the “statesmanship” of the current prime minister and the current leader of the opposition — and he/she will do so with pleasure and within seconds. Apparently, individuals are quite capable to make complex sets of values commensurable, even in situations where they may not be aware of the yardstick(s) they employ (if at all).

In chess, we tend to categorize this type of knowledge as “intuition,” in law we frame it as “opinion.” Opinions and intuitions are individual; we experience them as an inalienable part of our personalities, our identities even. Van den Herik’s Ph.D. thesis challenged De Groot’s opinion that intuition in chess cannot be modeled.<sup>23</sup>

Opinions do not have much scientific weight, at least at first sight. Nevertheless they are important for the sciences, as assumptions about their collective functionalities make out the foundations, where our scientific notions of welfare functions (the invisible hand is the result of aggregated opinions) and of democracy (as elections are a matter of aggregated opinions) are built upon. Opinions are special. They require individual autonomy, that is: freedom to act and responsibility for the consequences. In practice, opinions solve some serious problems for the social sciences. No social science has ever gotten very far on any path to predict individual preferential behavior.<sup>24</sup> This may be of some relevance if we accept that adjudication comes in the form of opinions. We (legal scientists) delegate those individual choices that add up to the level of fairness of our societies to those individuals we trust and we shield them with constitutional independence (i.e., freedom) to choose. I suggest that

- if economic science relies on the assumption that optimal welfare is functionally related to opinions (the freedom to choose in the market) and
- if political science relies on the assumption that optimal governance is functionally related to opinions (the democratic freedom to elect those that govern) it may well be that
- legal science relies on the assumption that optimal adjudication, as an important facilitator of a fair society, is functionally related to opinions (the constitutionally independent freedom of members of the judiciary to decide conflicts).

<sup>23</sup> At that time (1983) Van den Herik did not state that intuition can be modeled, but he stated (1) maybe it can be incorporated implicitly in a computer program, since it is inherent in the knowledge, or (2) maybe intuition is not a necessary ingredient for a computer program to play at world-championship level.

<sup>24</sup> Economists like Kenneth Arrow even get stuck in the observation that individual preferences show inconsistency.

To my own considerable amazement I have thus stumbled upon the observation that all three basic assumptions mentioned are the result of operational naturalistic fallacies: if what is optimal is considered good, while what is optimal is caused by factual choice, then all three assumptions mentioned include an is-ought derivation. Consequently, these assumptions are contestable as they employ the naturalistic fallacy. There is no room for pointing at the axiomatic status of opinions in our judicial arrangements and deny the deontic possibilities of computers because of the naturalistic fallacy implied. The axiomatic status of opinions in our existing legal arrangements proves to be very much contestable itself as it firmly rests on the same fallacy. My earlier misgivings about discussing dangerous questions are coming true, so it seems. Taking the naturalistic-fallacy argument serious implies that neither judges nor computers are fit to adjudicate.

I suggest facing the argument squarely and arguing the naturalistic-fallacy argument not to be a serious refutation at all. After all, it shows what we knew all along — that it may be true that we cannot derive what is right from what is. So what? The observation is hardly ever relevant as (outside science and mathematics) *derivation* is hardly ever possible or required. We need not take the naturalistic fallacy argument serious because it is beside the point. There is no or hardly ever any derivation involved in adjudication.

## 6 Conclusions

The argumentation leaves us with the conclusion that we have not found any compelling argument against the proposition that computers ought to adjudicate as soon as they perform consistently better than human judges. This result is not conclusive, of course, as there may be many other and possibly worthier refutations. It nevertheless is an attempt that does not weaken the proposition.

The foregoing discussion of the normative side of Van den Herik's proposition has brought two interesting issues to the fore, I think. One is the co-relation that seems to exist between the growth of knowledge and the diminishing of free space for rational choice. The second one is the observation that derivability is not required for quality adjudication. Both issues merit additional attention.

### 6.1 The Growth of Knowledge and the Loss of Freedom

I consider adjudication to be a professional activity. This implies, that it is performed by agents that know as much as can be known about their domain, yet need not know enough to claim that they are able to predict the outcome of any question within their domain. This implies that as more knowledge becomes available, fewer questions with unpredictable outcomes remain. I consider this correlation to be of importance when looking at adjudicatory trends. In general, I suggest that where the room for “free,” creative solutions becomes smaller, parts of the professional status (or glamour) may be influenced. There are several examples in adjudication.

- In the USA, in 1986, we have seen that the reduction of the free space for creative professional sentencing adjudication was diminished by the legislator, by law, introducing quite rigid sentencing guidelines. These guidelines were considered necessary when (statistic) information became available about the ways the judiciary employed their (former) freedom to adjudicate sentences for criminal acts. They consistently discriminated black criminals. As this was considered unacceptable, professional leeway was reduced drastically. This reduction of professional autonomy was motivated by distrust of (human) professional integrity.
- In the Netherlands, in the 1990s, the so-called Mulder Act changed professional freedom in a totally different manner: it was decided that the “lighter” traffic felonies no longer would be decided by judges, but would be decided in an automatic, administrative process. Thus reducing the adjudicatory domain diminished the professional freedom. The reason was, of course, that the amount of traffic felonies clogged up the legal system. This reduction shows that an economic argument concerning mass litigation may push the transition from human towards computer adjudication.
- In the Netherlands, in the 1980s, the continuous flood of cassation appeals on sentencing and sentencing motivation bothered the Supreme Court. A line of work was chosen, wherein these appeals were rejected in a very much-shortened procedure and without other motivation than a reference to established policy. This was considered to be against the requirement of decent feedback (motivation, explanation) in both the Dutch constitution and in the European Convention on Human Rights. Subsequently, a formal Act was passed, validating this practice. The reduction shows, that there may be some areas of professional freedom the profession itself does not want to keep and even manages to eliminate.

These examples suggest that there are at least three possible forces that reduce the professional freedom to adjudicate: (i) the way this freedom is exercised is considered below par by the legislator who applies its jurisdiction to reduce it, (ii) the sheer mass of litigations forces the adoption of uniform policies on economic grounds and (iii) the establishment of a fixed adjudicatory policy reduces the need (and motivation) for professional attention. All three forces work towards enlarging the possibilities for computers to adjudicate. As the reduction of professional adjudicatory freedom will reduce the authority of the established judiciary organization, it may well be that a balancing force can be found in the reluctance of the legal discipline to follow the path towards the growth of knowledge as referred to in footnote 3. Non-derivability supports the need for and the status of professionals. This may lead to the strange situation that *adjudicatory knowledge* will be acquired mainly by non-legal disciplines (like anthropology, sociology and economy). This may also explain why *adjudicatory practice* is becoming more and more the subject of serious competition by alternative (even online) conflict-resolution organizations.

## 6.2 Derivability and Quality Adjudication

The distinction of the adjudicatory domain into a predictable and an unpredictable sub-domain raises questions about the meaning of “quality adjudication.” Let us assume that the reciprocity equations mentioned earlier have been, are, and will be effective for the sub-domain that is predictable. Then, we may assume that those adjudications that in fact diverge from the predictable have low quality. These adjudications do occur under human adjudication. The Roger King case comes to mind, for example. Thus I tend to conclude that computers ought to adjudicate in predictable sub-domains. As of now.<sup>25</sup>

## 6.3 Afterthought

The foregoing shows the fate that hits every effort in Artificial Intelligence. As soon as tasks that require intelligence become within the reach of, say, intelligent agents, they seem to become mere Information Technology. A side effect is often, that AI efforts successfully expose some emperor’s new suit to be less glamorous than before. No reason for Artificial Intelligence to despair, though. As soon as it becomes acceptable that computers ought to adjudicate in predictable sub-domains, the question might be raised whether computers can legislate.

---

<sup>25</sup> The traditional argument against this opinion refers to all those seminal cases against established practice. An example would be the Dutch Supreme Court ruling of 1921, wherein was decided that the “theft” of electricity was possible, although the Dutch Criminal Code only considered theft of tangible objects a crime. I do not think much would have been lost when the Court had kept to the established policy. It would have induced the legislator to mend its ways and it would have prevented the Court to “legislate” in a non-democratic procedure.