Meta Ethics in Legal Reasoning

Hanno F. Kaiser[1] (2004)

1. Legal Reasoning

Legal judgments result from the application of rules (R) to facts (F). Rules are, at their most basic logical level, if-then conditionals. If certain elements (E) are fulfilled by the facts of the case (F), then the legal consequences (C), prescribed by the rule (R), apply. Suppose that R1 stands for the following rule: "Whoever steals apples (E) shall be banned from the store (C)." Suppose further that you have just been banned from the store, based on an application of that rule. If you disagree with the application of R1, you can argue your case (that is, deny the applicability of R1 to you) on four different grounds:

(i) The facts (F) are wrong ("It wasn't me.")

(ii) F is not a proper instance of E ("I didn't steal the apple, I merely exchanged it for a bad one that I had already purchased.")

(iii) R1 does not exist ("No one has ever heard of this store rule.")

(iv) R1 is invalid ("The store rule is unlawful/grossly unfair.")

Ground (i) is a factual argument, it can be resolved on empirical grounds, for example through witness testimony; (iii) is also a factual argument; it can be determined empirically whether the rule is a social reality in that it is being followed on a regular basis and/or that it has been properly promulgated. Grounds (ii) and (iv) are different. Whether exchanging an apple constitutes "stealing" or whether the store rule is invalid because it is unlawful or unjust can only be decided based on additional rules (R2) that govern the applicability of R1, for example, the canons of construction, or rules that govern the validity of R1, for example the common law of contracts or property. Of course, the applicability or validity of R2 is also open to debate, which requires

[1] Document version 1.2. Questions or comments? Email me at hanno@wobie.com.

recourse to even higher sets of norms (R3…Rn), until one arrives at the highest and most basic norm (Rb).[2]

In a legal system, Rb is a constitutional rule. However, the convention that Rb is, in fact, the highest norm (or, if that rule is part of the constitution, the extra-legal rule that the constitution ought to be followed) is itself a rule; thus, logically, the hierarchy of rules is open-ended.[3]

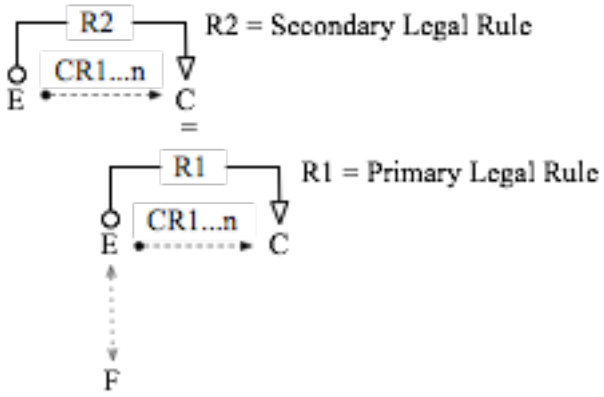
Legal systems avoid the impractical consequences of the infinite regress by continually invoking collateral rules from non-legal systems (CR), or by incorporating such rules into the law, to justify the validity of legal rules of any hierarchical level and to discourage parties from requiring proof of validity qua deduction from ever higher legal rules. The legal system imports collateral rules and principles from neighboring disciplines, for example, economics (efficiency), religion (marriage), psychology (feelings of inferiority because of unequal treatment), morals (abortion). The process of importing these rules is one of assimilation, that is, the legal system recreates originally external rules internally in the context of genuinely legal arguments, for example discussions of reasonableness, of due process, equal protection, common sense, fairness, equity, slippery slope and universalization arguments. ("What if everyone 'exchanged' their bad apples?"). As each of these collateral rules can be questioned in just the same way as rules of the legal system, bringing them into the mix does not solve the logical problem of the infinite regress; however, on a factual level, it very effectively deters the radical skeptic, as questioning not only the core legal rules that govern the conduct at issue but also the extra-legal or incorporated collateral rules would quickly exhaust the resources of any party and undermine the legitimacy of the protest against a specific rule by turning it into fundamental opposition to a broad set of rules and principles that most everybody (rightly or wrongly) subscribes to

---

[2] In our example R1 is a primary rule governing conduct. R2...Rn are examples of secondary rules that govern the applicability or validity of primary rules or lower-ranked secondary rules. Rb, the Basic Norm, the highest vantage point from within the legal system. See (Kelsen, 1997) and (Hart, 1997).

[3] Questions regarding the existence, applicability, and validity of legal norms are closely related to the definition of law. Virtually all such definitions refer to three elements: social efficacy (SE), proper promulgation (PP), and acceptable content (AC). Positivists and non-positivists both require a minimum of SE and PP to recognize a norm as a legal norm. In addition to SE and PP, non-positivists also include AC into the definition of a legal norm. Thus, for non-positivists there are extremely unjust norms (for example, laws that allow the indefinite detention of "enemy combatants" without a trial), that, even though they have been properly promulgated (PP) and are observed (SE) in practice, are not law for a lack of minimally acceptable moral content. Positivists would maintain that these rules are law, albeit morally corrupt law. See (Alexy, 1992) for one of the best discussions of these definitional issues of law.

without questioning. Justification, in practice, has therefore a vertical and a horizontal dimension. The chart below illustrates the rule-based character of legal reasoning.[4]



The sketch above illustrates not so much the futility of legal reasoning, but rather the importance of normative arguments. Normative arguments belong to a discourse about the validity of rules. In contrast, factual arguments, in the context of rules, belong to a discourse about the existence, that is, the facticity, of rules. Both discourses overlap in practice; a defendant is likely to make arguments (i)-(iv) ("if the law is bad, argue the facts; if the facts are bad, argue the law"); however, the criteria for what constitutes a good reason for the defendant's claims are different in the factual and the normative discourse. Factual claims are settled by true descriptive statements. For example, the statement "I have never been in the store" is true if, and only if, I have, in fact, never been to the store. The truth conditions of a descriptive statement can be established empirically.

It is less obvious how normative claims can be settled, for example: "R1 is invalid." Is this a question of fact, so that "R1 is invalid" is true if, and only if R1 is in fact invalid? If that's the case, what is the property (P) that valid rules possess and invalid rules do not? Can the presence or absence of P be detected empirically? If not, is there a special faculty (a "moral sense") that is required to intuit the presence or absence of P? If so, who possesses that faculty? Everyone or only those who have been properly trained and educated? What makes an education proper? To answer the

[4] The graph is a modified version of Toulmin's general structure of a rule based argument. In Toulmin's model, certain facts ("A was born in Bermuda.") are connected to a claim ("A is a British citizen.") by rules that serve as inference-warrants ("Whoever is born on British soil shall be a British citizen."). My modification is intended to highlight both the descriptive feature of the rule ("Whoever is born on British soil...") and its prescriptive properties ("... shall be a British citizen.") (Toulmin, 2003).

question how normative claims can be settled (that is, what passes as a good normative argument), we must first understand what normative claims are. That is the realm of meta-ethics.

2. The Linguistic Turn

Meta-ethics analyzes the language of morals, or the linguistic properties of moral arguments. The focus on language is useful, because language is the medium of moral arguments, if not of moral practice. Reasoning about morals is linguistic activity and governed by the rules of proper use of language. If moral words have special linguistic properties (for example, if "ought" implies universal applicability), such properties will guide our moral arguments and influence the results of rational discourse. The focus on language reminds us that we cannot start our reasoning with a clean slate; much of what counts as a good argument is determined by logic, and most of the world as we perceive it is experienced within a framework of deep grammatical structures (for example, that time is divided in past, present, future; or that our spatial coordinates - right, left, in front of, behind, etc. - are relative, not absolute) and of more volatile modules or plug-ins of beliefs and identities, that our various cultures and subcultures offer to us for subscription.

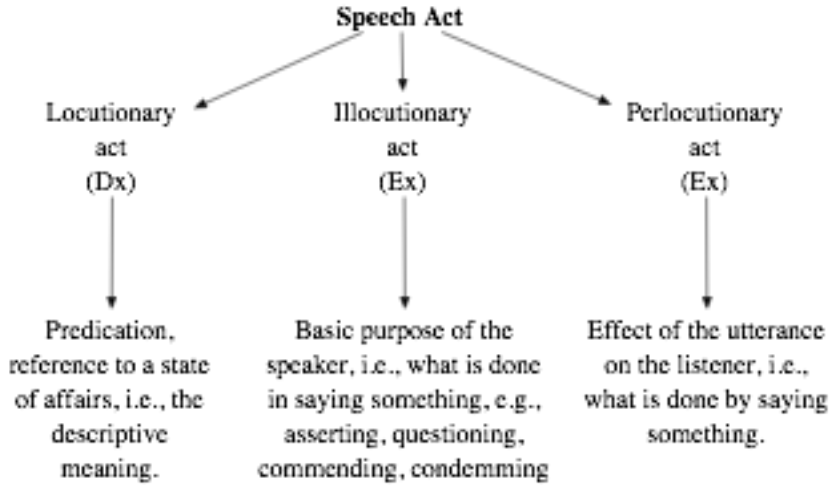
3. Meta-Ethical Positions

Given the inescapability of language for thinking about ethics, speech act theory has proven to be a useful tool for the analysis of normative claims. Utterances (for example, "the pill is red") have a range of effects (that is, they have multiple dimensions), including:

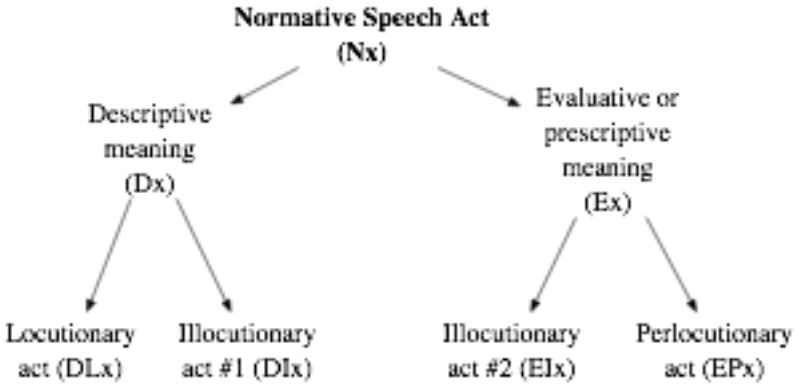
- The delivery of propositional content (here, that the pill is red) (locutionary act);
- The expression of the speaker's basic purpose and intended meaning (here, making a factual claim, that is, an assertion) (illocutionary act); and
- The effects of the utterance on a listener, which largely depend on context (here, for example, relief, if the listener is color-blind and thought that he just swallowed a poisonous blue pill by accident) (perlocutionary act).

For our purposes, it is useful to distinguish between the descriptive and the evaluative meaning of a normative claim. (Hare, 2000). Broadly speaking, the descriptive meaning (Dx) of the normative claim (Nx) "Jane did good, when she gave money to the poor" is the locutionary act, the delivery of propositional content ("Jane gave money to the poor"). The evaluative meaning (Ex) is what we do in making that claim,

which is commending Jane (illocutionary act). Thus, the meaning of Nx is a function of Dx and Ex, which can be written as Nx(Dx, Ex).



Note that this description of the linguistic properties of a normative claim is somewhat perfunctory. For a proposition to have truth conditions, it must be uttered in an assertive mood. Thus, Dx contains not only the locutionary act (DLx) but also the illocutionary act of asserting (DIx). The evaluative meaning (Ex) has been defined in terms of the illocutionary force of Nx (EIx); however, as we will see below, some authors have held that Ex should (also) be defined by the effects of the claim on others, that is, not by what we do intrinsically in making the claim (illocutionary act) but rather what we bring about extrinsically by making it (perlocutionary act) (EPx). Sometimes it is helpful to expand the simple concept of a normative claim Nx(Dx, Ex) to Nx(Dx(DLx, DIx)), Ex(EIx, EPx)), as summarized in the chart below.



a. Descriptivism

The great divide in the field of meta-ethics is that between descriptive and non-descriptive theories, where descriptive means entirely descriptive and non-descriptive

not entirely descriptive. (Hare, 2000). Descriptive theories claim that the meaning of Nx is determined solely by Dx, thus: Nx(Dx). Non-descriptive theories claim that the meaning of Dx is not determined solely by Dx; rather there is an additional evaluative element of meaning to a normative statement that cannot be reduced to a description of facts; thus: Nx(Dx, Ex). Descriptivism attempts to identify certain moral properties of acts or actors, for example "wrongness." If "wrongness" is present, then the act is bad.

- Objective naturalism claims that moral properties are ordinary public properties that can be perceived with our five senses (for example, the fact that the act promotes universal happiness, which could be measured by a survey);
- Subjective naturalism claims that moral properties are ordinary private properties that can be perceived through introspection (for example, the mental fact that I feel disgusted by someone's lies);
- Intuitivism in contrast to subjective naturalism claims that moral properties are of a non-natural variety. As they cannot be perceived with our five senses, detecting their presence requires a special private "moral sense" (for example, I intuit moral outrage when faced with someone's lies).

All descriptive theories translate moral words ("ought", "should", etc.) into descriptive statements. Moral claims are therefore factual claims (or they fully supervene upon factual claims), and as such, they have truth conditions. The claim: "It is wrong to harm animals for fun" is true if and only if it is in fact wrong to harm animals for fun, where "wrong" could be a shorthand for (i) the failure to promote universal happiness (utilitarianism); (ii) against the law (legal positivism); (iii) against God's will (religious positivism); (iv) lack of adaptivity (biological positivism); (v) repulsive (subjectivism or intuitionism); or any number of descriptive definitions of "wrong".

If finding universally applicable foundations for moral judgments is the goal, then the main problem with any form of descriptivism is that it collapses into relativism. This is rather obvious for subjective naturalism and intuitionism. Whether I correctly report my mental state of disgust (subjective naturalism) or feel morally outraged (intuitionism) may be subject to dispute (for example, I could be lying); however, if A and B are both truthful and A feels outraged and B doesn't, that is the end of the discussion. There is no real disagreement; A reports her mental state or intuition and B reports his. Once we start adding rules to distinguish relevant intuitions from irrelevant ones (for example, we could stipulate that only the intuitions of men, or of the well-educated should count), we are replacing private truth criteria with public ones and are thus converting subjective naturalism or intuitionism into objective naturalism.

Objective naturalism captures an important insight about the moral discourse, which is that people do, in fact, disagree – often vehemently so.[5] Public truth criteria for Nx account for the possibility of disagreement. If A claims that C's actions were permissible and B claims that C's actions were impermissible, A and B are contradicting each other as opposed to merely reporting incompatible states of mind. In the case of genuine normative disagreement, A and B cannot both be right whereas in the case of contrary reports of incompatible states of mind they could. Thus, it is a condition for the possibility of genuine normative disagreement that the criteria for determining who is right are in some sense public. Various empirical criteria meet the publicity requirement, for example whether higher rules exist (legal positivism) or whether the act in question, in fact, "maximizes the satisfaction, in sum, of the preferences of all affected parties" (utilitarianism). (Hare, 2000). Unfortunately, empirical public truth conditions not only make normative disagreement (and agreement) possible, they are plagued by widespread and persistent actual disagreement. Where, over time, many descriptive discourses have converged towards a consensus-equilibrium, at least for certain periods of time, the same is not true (any more?) for normative discourses. It might be possible to empirically demonstrate overwhelming support for Nx in a certain sub-community (for example, "abortion should be illegal"); however, there are other sub-communities within which that demonstration would most certainly fail. Which communities' standards should govern? Even with less controversial claims (for example, "killing prisoners of war is morally bad"), the problem persists. While support for the less controversial claims would be quantitatively greater, even supposing that support would approach de facto consensus, a dissenter could still make a coherent counter-argument. Ultimately, empirical truth-criteria are conventional, which is why objective naturalism too collapses into relativism. That is not a flaw of

[5] Note that the objectivity of the moral judgment clashes with the intuition that correct moral judgments are practical in that they translate into reasons for actions. If I say "I am morally obligated to give to the poor" and then ignore the beggar in the street, holding in my hand a quarter that I don't need, people would find my behavior inconsistent and perplexing, because my moral judgment apparently failed to motivate me. (Most likely, one would question the sincerity of my beliefs.) The problem with the intuition that moral judgments are both objective and practical is that "unfortunately, [the metaphysical and psychological] implications [of that intuition] are the exact opposite of each other." (Smith, 1994). Per the standard model of human psychology (which we owe to D. Hume and, more fundamentally, to Plato's distinction between reason and experience as sources of knowledge), there are two mutually exclusive psychological states, beliefs and desires. Beliefs may be true or false but they don't motivate. Desires motivate but they are neither true nor false. The implications of the objectivity of the moral judgment are moral realism and cognitivism; the implications of the practicality of the moral judgment are irrealism and non-cognitivism. Whether that result is truly a dilemma depends on the continued viability of Hume's model of human psychology.

the theory if descriptivism is understood as an explanatory theory. However, once descriptivism is used to justify prescriptive claims (for example, "You should do x because it promotes universal happiness") it overextends itself. Any descriptivist's claim is subject to Moore's open question argument, because it is always possible to ask: "Granted, x promotes universal happiness (or whatever the definition of 'ought' may be), but is it right?" The table below gives an overview of some of the structural elements of the most important variants of descriptivist ethical theories.

Theory Nx(Dx)	Properties (Dx)	Publicity (Dx)
Objective Naturalism	natural	public
Subjective Naturalism	natural	private
Intuitionism	non-natural	private

b. Non-Descriptivism

Imperative Theory

The defining feature of descriptivism, the translation of moral words into wholly descriptive statements, that is, Nx(Dx), came under attack by what has been labeled "imperative theory". According to imperative theory, normative claims cannot be translated into descriptive statements. Rather, normative claims (such as, "stealing is wrong") are, in fact, imperatives in disguise ("Don't steal!"). Imperatives have certain linguistic properties, that are entirely unlike those of factual statements, most significantly that of a "verbal shove". (Hare, 2000). A speaker uses normative claims (imperatives) to get someone to do something. Imperative theory thus negates descriptivism in toto and claims that the meaning of a normative statement is solely determined by its evaluative meaning Nx(Ex) and that the evaluative meaning takes the grammatical form of an imperative.

There are several serious problems with that position. First, while it is certainly possible to derive imperatives from normative claims, there is no necessary connection between the two, as there are many imperatives without corresponding normative

claims. (What is the normative claim underlying a command to shut the door?) Worse, imperatives obscure the defining feature of normative claims, which is their necessary connection to or their expression of an underlying rule, a feature that imperatives do not share, as it is perfectly possible to issue entirely arbitrary commands. Second, imperative theory leads straight into irrationalism, as imperatives have no truth conditions whatsoever. The command "Shoot the prisoners!" is neither true nor false. The only thing that can be said about its facticity (other than whether it has been uttered or not) is whether it has been followed or not. Third, and closely related to the second point, the "verbal shove" theory identifies the evaluative meaning (Ex) of an imperative with the perlocutionary effect of the utterance. The speaker uses a command to get someone to do something, consequently, the meaning of an imperative is determined by its effects on the addressee. Relying on perlocutionary effects to determine the meaning of an utterance has proven to be difficult if not impossible, because the effect on the addressee depends on the internal mental states of the addressee. As humans are "non-trivial machines," our reactions to symbolic inputs are highly contingent.[6]

Ultimately, it is a factual question whether patterns of behavior emerge in reaction to commands so that a conventional meaning may be established. (In certain environments, that is clearly the case, for example, in the military or in squad-based computer games.) However, for the class of imperatives that we are concerned with here (that is, those whose justifiability is contested), the unpredictability of their effects on others is too great to permit the gradual formation of a stable meaning. Thus, if we take imperative theory seriously, Nx(Ex) has no descriptive truth conditions (because there is no Dx) and no discernible internal logic (Ex). Consequently, normative statements, Nx(Ex), cannot be discussed rationally.

Emotivism

The failure of imperative theory's radical attack on descriptivism gave rise to emotivism, which is related to imperative theory in that it defines Ex though the perlocutionary effects of a normative statement, but differs sharply from imperative theory in that it includes (usually public) descriptive elements (Dx) in the definition of a normative claim (Nx). Emotivism gave normative claims their modern conceptual form:

[6] Trivial machines, no matter how complicated, can be understood in terms of input and output, ex post causation, and means-ends rationality, in other words, they are synthetically determined, independent of the past, analytically determinable, and predictable. Non-trivial machines, in contrast, cannot be understood in that manner; their behavior depends on changing inner states and continuous self-reference. Non-trivial machines are historically dependent and unpredictable. (Foerster, 2002).

Nx(Dx, Ex). The inclusion of public descriptive elements (Dx) allowed emotivism to account for genuine normative disagreement (of sorts), where A and B agree on Dx (for example, that abortion kills a fetus) but attach different evaluative meaning (Ex) to the facts. The qualifier "of sorts" is justified, because Ex is defined by the perlocutionary effect of Nx. Thus, if A opposes abortion and B supports it, A is saying: "Agree with me [verbal shove] that abortion should be illegal," and B is saying: "Agree with me [verbal shove] that abortion should be legal." While A's and B's policies are different (they both want the addressee of the utterances to do contradictory things), their statements are only contradictory in the derived sense that the same addressee cannot consistently choose to agree with both A and B. As Ex is defined by Nx's perlocutionary effect, emotivism remains open to the charge of irrationalism.

Universal Prescriptivism

Enter rationalism, which, across its various forms, agrees with emotivism in that the concept of a normative statement entails both (usually public) descriptive (Dx) and evaluative (Ex) elements, Nx(Dx, Ex). However, unlike emotivism, rationalism defines Ex through the illocutionary meaning of Nx. Unlike the perlocutionary effects, the illocutionary meaning of an utterance is usually not subject to dispute, even if the perlocutionary effects of that utterance may be highly contingent. For example, if speaker S says to addressee A: "Abortion should be illegal," the perlocutionary effect on A is uncertain, as A's reaction to S's claim might range from assent to violent disagreement. The illocutionary meaning, however, is clear; S condemns abortion. Thus, there exists a sufficiently stable basis for the construction of a conventional logic of normative claims. That logic can be used to distinguish normative claims that are formally correct from those that are formally incorrect. Irrespective of their substantive content, the latter ones are bad arguments.

The next step is critical, because it is here where the logic of the ethical discourse blends into its substance. Universal prescriptivists, most notably R. M. Hare, claim that moral words (for example, ought), imply a logic of universalization. (Hare, 2000). If under circumstances D1…Dn A ought to do x, then under the exact same circumstances (which include critical features of A, for example the fact that he is a policeman) everyone ought to do x. Hare then uses the universalization feature to connect the implications of Nx to the preferences of the speaker. For example, if A were to claim that "under circumstances D1…Dn, torturing for fun is permissible," A would have to accept that, under the exact same circumstances, everyone else may be tortured for fun, including A. While in practice, A might be a moral risk-taker who considers the chances of being in the position of the victim as sufficiently slim and who therefore includes the risk of being tortured in his or her own preferences, Hare

redefines what the "preferences of the speaker" are. They are not, in fact, the preferences of the speaker but rather the preferences of the hypothetical victim; therefore, the so-defined speaker A "cannot will" (in a Kantian sense) what the victim, in fact, does not will.[7] (Hare, 2000). If nothing else, the requirement that the speaker adopt the victim's preferences wholesale is not dictated by the logic of the moral words; rather, it is a substantive moral claim of equality, pursuant to which the preferences of everyone who is affected are entitled to at least some weight. The universalization feature is one of logic, but the "changing-places" or golden rule argument is not.

Transcendental Pragmatism

Another blend of rationalism is transcendental pragmatism (or discourse theory), which incorporates not only descriptive elements (Dx) and evaluative meaning (Ex) defined qua illocutionary force into the meaning of Nx, but also the necessary (and usually implicit) assumptions that a speaker must make to claim or to dispute Nx in earnest. (Apel, 1999). The assumptions (A) are not part of the premises of the normative argument, rather, they are necessary conditions of the act of arguing. For example, the claim "I don't exist," taken at face value, is a performative contradiction because I must exist to make the claim. With respect to normative discourse, the main contribution of transcendental pragmatism is to expose certain persistent challenges to normative claims as performatively flawed. Flawed challenges neither warrant, require, nor permit refutation. For example, in the context of a discussion about rationalism, a sceptic might say: "Why should I be rational?" A transcendental pragmatist would claim that this is a defective question, because in questioning rationality, the sceptic implicitly assumes that there could be a rational answer to her question, that I (as the addressee of her question) can give that answer, and that she can understand my answer, which, if convincing, could make her change her mind. If the sceptic makes no such assumptions, then she is not *really* asking a question, in which case I am absolved from having to answer. The central theme of transcendental pragmatism is thus to add the conditions of arguing to the logical requirements of the argument. The inclusion of the necessary performative assumptions of questioning Nx is a perlocutionary effect of sorts, because it refers to what the speaker does by making the utterance, which is, to reveal certain assumptions (that is, claims) that the speaker implicitly assumes to be true. If Nx contradicts the claims implicit in these assumptions, then the speaker contradicts herself, and the apparent question should be disregarded as faux skepticism.

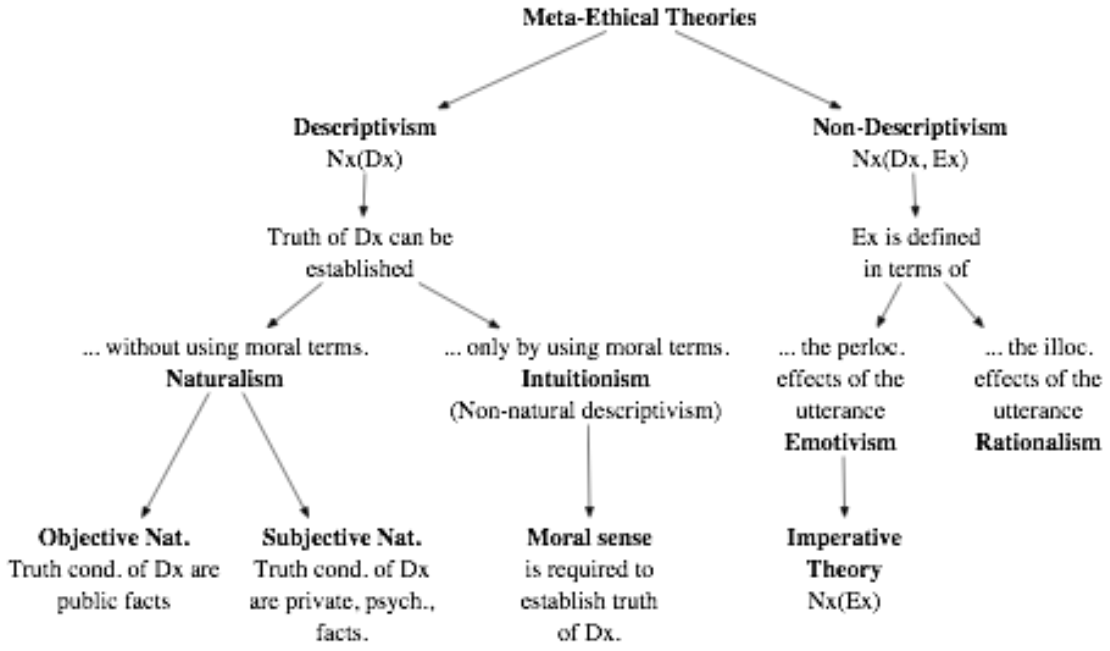
[7] Hare presents this argument in the reverse; first, he establishes what I cannot will, then he concludes that what I cannot will cannot logically be willed as a universal maxim.

Some defining features of non-descriptivist theories are summarized in the table below:

Theory	Structure	Ex	Authors
Imperative Theory	Nx(Ex)	Perlocutionary Act	Austin
Emotivism	Nx(Dx,Ex)	Illocutionary Act	Stevenson
Rationalism	Nx(Dx,Ex)	Illocutionary Act	Kant, Hare
Transcendental Pragmatism	Nx(Dx,Ex)	Illoc. and perloc. Act	Apel

c. Summary

The chart below summarizes the meta-ethical positions discussed in this article:



Selected Bibliography

Alexy, R. (1992) Begriff und Geltung des Rechts. Alber, Frbg.

Apel, K.-O. & Papastephanou, M. (1999) From A Transcendental-Semiotic Point of View. Manchester University Press.

Foerster, H.V. (2002) Understanding Understanding: Essays on Cybernetics and Cognition. Springer Verlag.

Hare, R.M. (2000) Sorting Out Ethics. Oxford University Press.

Hart, H.L.A. (1997) The Concept of Law (Clarendon Law Series). Oxford University Press/

Kelsen, H., Paulson, B.L. & Paulson, S.L. (1997) Introduction to the Problems of Legal Theory: A Translation of the First Edition of the Reine Rechtslehre or Pure Theory of Law. Clarendon Press.

Smith, M. (1994) The Moral Problem (Philosophical Theory). Blackwell Publishers.

Toulmin, S.E. (2003) The Uses of Argument. Cambridge University Press.

***