SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #01. April 1995. Pp. 13-45.
Natural Kinds and Projectible Predicates

NATURAL KINDS AND PROJECTIBLE PREDICATES[1]

Axel Mueller

1.— Introductory Remarks

In this essay I want to approach two — at first sight not immediately connected — themes:

1.) Goodman's Paradox, i.e. a problem usually associated with the justification of induction or the conditions of confirmability of hypotheses, and

2.) some traits of the application of the so-called «natural kind terms» as they have been postulated by proponents of the theory of direct reference, i.e. theses and problems usually associated with the interpretation of possible world discourse and/or metaphysical questions as to «metaphysical realism» and essentialism.

Do these two problem clusters intersect in any sense at all? One intention of the following reflexions consists in an attempt to answer positively to this question. This might not seem to much a dare, as Goodman himself pointed out the connection between counterfactual conditionals, lawlikeness of generalizations and the problem of the characterization of projectible predicates, as well as Putnam always insisted in the «theoreticity» of natural kind terms, that is, understood them in the sense of the predicates which are used with more or less success in confirmation — and induction-impregnated practices. Nevertheless there is little more than hints in the respective direction from either side. So Goodman says that to entrench a «class of objects» and to entrench a predicate is more or less the same[2] and adds, in the part with the title «Survey and speculations»: «Our treatment of projectibility (...) may give us a way of distinguishing 'genuine' from merely 'artificial' kinds (...) and thus enable us to interpret ordinary statements affirming that certain things are or are not of the same kind (...). [S]urely the entrenchment of classes is some measure for their

[1] Final form of this paper has benefited very much from the minutious reading and discussing of a first draft done by the editor of this journal, Lorenzo Peña, as well as from a discussion group in the Spanish Institute for Advanced Studies (CSIC), Madrid (Spain), formed by Carlos Thiebaut, Cristina Lafont and Lorenzo Peña. Thanks to all of them. Working extensively on this paper, which is a draft of the line of argument of my dissertation, was possible thanks to a grant of the Studienstiftung des deutschen Volkes.

[2] Cf. Goodman, N.: *Fact, Fiction, and Forecast* [in the following FFF], Hassocks

genuineness as kinds; (...) An adequate theory of kinds should in turn throw light on some troublesome questions concerning the simplicity of ideas, laws and theories.» (p.122-3)

Putnam says that to stay with a predicate and to treat two theories with different characterizations of its reference-class as successors, i.e. phases of one and the same global theory, is virtually the same[3]. On the other hand there has been a considerable progress in the theory of reference concerning natural kind terms, which has not yet had its due resonance in confirmation-theory[4]. Two contingent historic facts might have prompted this situation: first there is the unhappy divorce of epistemology and metaphysics and the subsequent dismissal of epistemological concerns promoted by Kripke and the theorists of direct reference mainly interested in ontlogical questions. On the other hand we have the implicit or explicit assumption of the unintelligibility of possible world discourse as «intensional» and the subsequent assumption of insignificance concerning the results of «natural kind term theory» of theorists of science interested in questions of confirmation theory. My impression is that both a priori rebuttals are unjustified. One need not accept the Kripkean essentialistic self-interpretation of reference theory (with natural kinds as real essences which dictate us what ontological commitments to make, assuming the truth of our theories) to accept its pragmatic and normative, as well as its purely linguistic imports[5]. And one need not become an ontological or epistemological sceptic when one accepts the

[3] p.95: «The entrenchment of a predicate results from the actual projection not merely of that predicate alone but also of all predicates coextensive with it. In a sense, not the *word* itself but *the class it selects* is what becomes entrenched, and to speak of the entrenchment of a predicate is to speak elliptically of the entrenchment of the extension [=reference, A.M.] of that predicate.»

[4] Exceptions to this can be found in the works of J.Leplin concerning his concept of «methodological realism» (see fn47) and S.Blackburn *Reason and Prediction*, Cambridge MA 1973, ch 4, who gives a realist account of Goodman's paradox.

[5] This has been demonstrated by interpretations of this theory given by H.K.Wettstein «Demonstrative Reference and Definite Descriptions» in: *Philosophical Studies* 40 (1981), 241-57, «Has Semantics Rested on A Mistake?», in: *Journal of Philosophy* 83 (1986), 185-209; «Cognitive Significance Without Cognitive Content», in: Almog, J. &al. (eds.): *Themes from Kaplan*, N.Y. 1989, 421-454, «Turning the Tables on Frege or How is it That «Hesperus is Hesperus» is Trivial?», in: Tomberlin, J.E. (ed.): *Philosophical Perspectives* 3: *Philosophy of Mind and Action Theory*, Atascadero (Cal.) 1989, 317-39, and N.U.Salmon («How *Not* to Derive Essentialism From the Theory of Reference», in: *Journal of Philosophy* 76 (1979), S. 703-725, as well as *Reference and Essence*, Princeton 1981 and «Reference and Information Content: Names and Descriptions», in: Gabbay, D./Guenthner, F. (Eds.): *Handbook of Philosophical Logic*, Vol. IV: *Topics in the Philosophy of Language*, Dordrecht 1987.

deepness of the problems of underdetermination raised by the discussions in confirmation theory by philosophers like Goodman and Quine[6].

To avoid these consequences and to keep the respective theories might seem to most of the philosophers of either part tantamount to drop the theory: direct reference without essences and necessary truth is like underdetermination without ontological relativity and incommensurability, as it were.

But there are always other possibilities apart from dogmatism.

There is, for example, a quite modest, pragmatic hypothesis which has been put forward by philosophers like Dagfinn Føllesdal and Keith S. Donnellan since the sixties, and there are Putnam's attempts to combine a critical epistemological attitude with a pragmatically biased modest realism stemming from or localizable in certain reference-theoretic assumptions. My attempt in this paper is to contribute some more programmatic considerations to this program. The basic idea consists in taking the theory of the direct reference of natural kind terms as an answer to the problems raised by the radicalization of underdetermination. In Putnam's case this switch from scepticism as to reference to an argument very much like 'if (1) there is no principled way to reduce the meaning to any epistemologically priviledged basis, (2) meaning is a matter of intratheoretical structure (interrelations of signs) and (3) meaning should determine reference, then non-(3) meaning does not determine reference, thus (4) reference being relatively independent from intratheoretical «meaning», so we have to provide an alternative account of reference' is evident. In this argument, as we see, there is no refusal of underdetrmination: (1) is entirely accepted. Neither is this possibilitated by a new foundationalism: (2) is accepted, thus (4) does not mean that reference is entirely «theory unloaded», i.e. independent of any theory, but there is no one theory which (now or in the future or in a world described by «necessary truths») determines reference. Reference is thus rescued to be the complicated thing it is: as the concept which serves to explain the relation between theory, understanding and the objects described, and is not determined by anything, factual or counterfactual, without reflexion on side of the users of theory. It is, in other words *supposed*. On the other hand, (2) prevents us from becoming Milleans and divorce theoretical terms from our understanding of them and their place in theories: intratheoretical reduction and definition is thus vindicated as a legitimate possibility, so that there are no grounds to suspect that what is being worked out is something like the «furniture of the world». (3) is, after all, quite a modest modification (although it goes right to the heart, one should add).

Taking this as an example, in the following I want to adventure the following ideas:

The conditions for and presuppositions (or commitments) of the adequate use of empirically interpreted predicates made explicit in the theory of the reference of natural kind terms coincide largely with the desiderata for a solution of Goodman's paradox[7]. I assume, in other words, that the referential anomalies resulting from

[6] The chief example of this attitude seems to be Putnam, although he as well should count as one among the theorists named before.

[7] This is to say, I haste to add, that I neither pretend to give a solution (because there is none) nor to abund in the theory of identity in modal logic.

«intensionalism» detected by Kripke, Donnellan and Putnam are *not only* and *not unseparably* such of the interpretation of formulae of modal logic, and that the Goodmanian anomaly ist *not only* one within the framework of confirmation theory and the theory of valid inductive method. In contrast to that I would propose to see both of the mentioned disciplines as «*contexts of discovery*» of one or more underlying, principal problem(s) for the philosophy of language as such which challenges certain ways of transforming old philosophical problems in problems of the philosophy of language. Thus I think that the metaphysical problems stemming from the discussions in the theory of direct reference are reinterpretable (even if this might be exactly what their proponents do *not* wish to do) as parts of answers or proposals for understanding the (normative-apriori) conditions for the justification and «normal» application of predicates within inductive practices which we always have to buy if we do use them in the «normal» way, i.e. assume inductive validity for our inferences from data. That is: they may be «internalized» and be seen as a description of the realism which guides us as long as we *use* the terms. On the other hand, Goodman's paradox might be seen, as I think, as a critical obstacle to a metaphysical hypostatization of *the* world, i.e. to the reification of something normative which is operative *within* our practices: it shows that we, as soon as we reflect upon these conditions, get to see that they always could be otherwise and that there is no ontological or otherwise *guarantee* for the correction of our conceptual schemes. We have to be realists to pursue the aims of science but we are not damned to live in one specific world and could not be so.

In short: I want to argue for a «deflationist» reading of the theory of direct reference combined with an «inflationist» reading of Quine-type (or, in general: instrumentalist) scepticism concerning the ontological import of theoretical concepts respectively the epistemological importance of the theory of direct reference. I understand this as a part of the elaboration of a concept of «world» or «reality» which helps us understand the rationality incorporated in the methodology of certain enterprises, like science. Thus both modifications could, at least as I hope, contribute to an elucidation of the ontological and epistemological premises which are operative in our use of language with empirical import[8].

These are two defects which I want to be clear about from the beginning; they are due to the general character of the theses I want to put forward: they should be valid, I think, for every account of identity through possible worlds, because they do not concern the concrete structure of
an assumption of sammeness of kind as such but its place and unavoidability in certain practices. I suppose that the most natural reading of the following results from the assumption of a modified Kripke-semantics for possible world like the one proposed by Deutsch in «Semantics for Natural Kind Terms», in: *Canadian Journal of Philosophy* 23/3 (1993), 389-412, and his improvements in «Semantic Analysis of Natural Kind Terms», in: *Topoi* 13 (1994), 25-30. However, as I said, the concern of this paper is less in semantics proper than in pragmatics.

[8] We can find witnesses for this suspicion on both sides. Thus H.K.Wettstein thinks that you simply miss the point of the theory of direct reference if you look for it exclusively in its aptness to formalize metaphysical speculation or in its contributions to the clarification of the interpretation of

The first part of my thesis is that one can obtain the most important results independent from the presupposition of a metaphysical realist interpretation of the modalities because modalities are not all that matters to epistemological matters, as all the world agrees. On the other hand, and this is the second part of the thesis, a pragmatic interpretation of the structural properties assigned by this theory to the use of empirically (or otherwise objectually) interpreted general terms can provide us with a non-naturalistic description of the characteristics of a possibility to use language which is of priviledged importance in contexts where we are primarily interested in learning from experience.

2.— Aspects of the theory of reference for natural kind terms: some remarks on the conditions for a distinction between «normal» general terms and «natural kind terms»

If one views the reference of a descriptive general term as given by a necessary and sufficient condition of its application stated in other terms than the general term in question (i.e., normally a description), there is room for a conflict between the satisfaction conditions associated with the condition for application and the reference of the term interpreted through it. In certain contexts both seem plausibly to be not completely substitutable. Thus if you determine the reference of the term «gold» with a description of the form (1) «something is gold iff it is F, G and H» and affirm (2) «It is possible that gold is not F» (e.g. on aposteriori grounds or in a thought experiment) then you get by substitution the inconsistent result that (3) «It is possible that what is F, G and H is not F». Nevertheless it does not seem that by your modal remark you construe any impossible or grammatically or logically false nor absurd affirmation. This would be trivially the case, of course, if you view (1) as a definition in the strictest sense of the word. In that case eliminability is carried through in virtue of the fact that (1) is an adequate definition (i.e. provides eliminability and non-creativity in the language where it occurs and is held true), and consequently (2) is inadmissible in a language where (1) is true. So avoiding (3) is possible by adopting an aprioristic point of view concerning the descriptively fixed reference which immunizes (1) from revision by hypotheses like (2), confirmed as they might seem. This is, however, an epistemologically quite uninteresting case. The interesting case is the one where you propose a *revision* or *alternative* to affirmations like (1) on whatever grounds, i.e. when you want to (and, strictly: have to) appeal to something like (2) to inspire an investigation as to whether (1) is true or not. This is what a change in status from a definition to a hypothesis seems to consist in, and one necessary step in this course seems to be exactly to *admit* (2), be the

modal discourse. In «Turning the tables on Frege or How is it that «Hesperus is Hesperus» is trivial» he expresses this view as follows: «If one sees the modal arguments as at the core of the anti-Fregean approach, as I do not, one might conclude that intellectually mediated reference [i.e. the determination of extension by intension, A.M.] is *not* what the anti-Fregean revolution is about» (p.336, my italics), but, as we could add, in the theory of interpretation for modal logic. In
«Cognitive Significance without Cognitive Content» (in: Almog, J./Perry, J./ Wettstein, H. (eds.): *Themes from Kaplan*, N.Y./Oxford 1989, 421-54) he considers to be the «lesson of the anti-Fregean revolution» the insight that «linguistic contact with things —reference, that is— does not presuppose epistemic contact with them» (454).

specification of «gold» what may. Another important thing seems to be that the admission of (2) goes *without*, from the point of view of the possibility of interpretation of the term, causing a complete deviance from former use or the inacceptability of a theoretical system which would inevitably prompted by such a patent contradiction like (3). A criterion for holding on to «former use» is beyond doubt to carry on the reference of a term. So the aprioristic attitude towards assumptions like (1) does not seem adequate for cases like the evaluation of hypotheses and the consideration of alternatives.

In the sixties thinkers like Dagfinn Føllesdal and Saul Kripke (among others) began to view this kind of problem as a symptom for an at least incomplete conception of the reference of descriptive terms and their behaviour in all contexts. They proposed instead to interpret the modal operators as relative to certain fixations of the reference of the non-logical terms of the languages in question. The central idea in these approaches seems to be a radical change in the conception of the status of sentences like (1). To introduce, use and learn some descriptive term usable in the above mentioned contexts (a «genuine singular term»[9] or «rigid designator»[10]) one fixes in a certain manner (operationally, ostensively, contextually or even with the help of a theory) its reference by the use of an implicit or explicit description, but this specific manner of making someone familiar with the reference of a term is neither to be seen as *a priori* successful in all possible circumstances nor necessarily true nor obligatory («analytic» or «true by definition»)[11]. On the contrary, whatever

[9] This is Føllesdal's term who introduced it in his dissertation *Referential Opacity and Modal Logic* (Harvard 1961) and explained its use further in the articles «Quantification into Causal Contexts», in: Cohen/Wartofsky (eds.): *Boston Studies in the Philosophy of Science*, Bd. II, N.Y. 1965, 263-74, reappeared in: Linsky, L. (ed.): *Reference and Modality*, Oxford 1971, 52-62, «Knowledge, Identity and Existence», in: *Theoria* 33 (1967), 1-27, «Interpretation of Quantifiers», in: Rootselaar, B. van/Staal, J.F. (eds.): *Logic, Methodology and Philosophy of Science*, Amsterdam 1968, 271-81, «Quine on modality», in: Davidson, D./Hintikka, J.(eds.): *Words and Objections: Essays in Honour of W.V. Quine*, Dordrecht 1968, 147-57, «Situation Semantics and the 'Slingshot' Argument», in: *Erkenntnis* 19 (1983), 91-8, «Essentialism and Reference», in: Hahn, L.E./Schilpp, P.A. (eds.): *The Philosophy of W.V. Quine*, LaSalle 1985, 97-113.

[10] This is, as everybody knows, Saul Kripke's term, who explained it and the premises for its application mainly in «Identity and Necessity» (in: Munitz, M. (ed.): *Identity and Individuation*, N.Y. 1971, S.135-64) and «Naming and Necessity» (mit Addenda) (in: Harman, G./Davidson, D. (eds.): *Semantics of Natural Language*, Dordrecht 1972, S.253-355 bzw. S.764-9).

[11] In a certain sense one can see this, at least in Føllesdal's case as a consequent application of Quine's critique of the analytic-synthetic distinction like the one pronounced in «Carnap and Logical Truth» (in: Hahn, L.E./Schilpp, P.A. (eds.): *The Philosophy of Rudolf Carnap*, LaSalle 1963, pp.385-406) where he says about definitions, which he cosiders to be the candidate of whose analysis we can most probably hope to get a notion on analyticity which does not coincide with logical truth: «Definitions (...) can be

means you choose or however you try to introduce a term with a fixed reference to someone or in a specific context, this can only be successful if you manage by this to get the reference *of the term* right, i.e. possibilitate that it be employed furtheron to refer to *relevantly* «the same» objects or, in case there are none of these, to none[12].

The point of «genuine names» is that they neither are implicitly nor imply any specific description to be used correctly. At least they do not have to be interpreted thus, in contrast to «usual» terms. So what has to be done is to find means to draw a distinction between «genuine singular terms» and disguised descriptions to be true to their respective differences in behaviour under certain interpretative circumstances and to avoid inconsistencies. For the satisfaction of the truth conditions of descriptions in different possible worlds coincides most probably, if these worlds differ substantially concerning the intended domain of the term, with a variance of its extension. Now, if «fixed use» coincided with «complee extensional determination», then it should be expected that a term whose reference has only been fixed for a *part* (e.g. «the thing *in the actual world*») of the «absolute» extension (through possible worlds and all times) either would be hopelessly unclear in its use or, if this is not accepted, as uniquely referring only to this factual, *partial extension*

either legislative or discursive in their inception. But this difference is in practice left unindicated, and wisely; for it is a distinction between particular acts of definition (...) So conceived, conventionality is a passing trait, significant on the moving front of science but useless in classifying the sentences behind the lines. It is a trait of events and not of sentences.» (p.395)

[12] For reasons that, as I hope, will become clear in the following, I depart here to a certain extent from the «orthodoxy» of direct reference theory, because I want to make a more general use of its results without an essentialistic commitment from the outset. This is why I do not refer to «microstructures» or «object-identity» but rather introduce contextually an unspecified notion of «relevant sameness» which is evidently much broader than e.g. Putnam's «same» (sc. «The Meaning of 'Meaning'», in: Putnam, H.: *Philosophical Papers 2. Mind, Language and Reality*, Cambridge MA, 1975, pp.215-71) or most of the other conceptions which have been developed in the framework of this theory (e.g. the writings of Salmon, Deutsch mentioned above). I consider it sufficient for the following to suppose some «sameness-in-use-relation» accepted by the users of singular or general terms in certain practices which are linked to inductive method and hypothetical reasoning. Each of these practices, as well as each discipline, will have its own specification of this relation of the form: «A is the same substance as B iff ...», «A is the same (historical,...) individual as B iff ...» etc., where «...» is probably an interpretation-condition drawing on admissible model-classes ('physically', 'chemically', 'historically' or otherwise admissible). Thus I am not necessarily referring only to «rigid designators» in the classical sense of unqualified identity, but to designators which are to be understood as rigid *within* each admissible model class. In that sense, substitutivity or identity seems to me to be a structure to be *aimed at* in the (a priori) evaluation of admissibility but not to be *ontologically presupposed*.

(which would then be its *full* extension, other intuitively acceptable applications out of this factual extension, if seen as correct, automatically provoking the assumption of a homonym but different term, a lexical variant). Both possibilities seem to be highly inadequate if we look to our actual behaviour when we use e.g. empirically interpreted terms and extend or differentiate their use: there is a lot of discount of differences in belief, as Putnam frequently says, and, above all, no assumption at all of a «change in reference». Føllesdal[13] expresses this change of perspective quite decidedly when he remarks concerning the function of concept-explanations («senses»): «genuine singular terms have a *sense* (...), and (...) they refer partly in virtue of this sense. However, while Frege held that sense determines reference (...) I hold thet reference «determines» sense, *not by itself*, but in an interplay with our theories of the world and our conception of how we gain knowledge and how we are likely to go wrong in our perception and in our reasoning. (...) *The sense of a genuine singular term is designed to insure through the vicissitudes of increased insight and changing scientific theories that the term keeps on referring to what it presently refers to.*» (p.112)

Thus the conclusion was, that «genuine names» should refer in all possible worlds to the same object. As this now has been assumed not to be automatically accomplished by appeal to some (criterially understood) descriptive condition (like the mechanisms envisaged in formulations as «in a purely semantic way» or «by the meaning of the terms»), this demand for «referential transparence» can be seen as an at least partly independent claim in its own right about the behaviour and use of empirically interpreted concepts. Further, as this is obviously the consequence of a general, metatheoretical reflexion on the status and possible function of «meanings», the same is valid as much for singular as for general terms. If there is any justified doubt as to how «referential transparence» is to be understood theoretically, then this cannot only affect a certain kind of terms (although it might be of heuristic importance to isolate the most evident case, as is the case of proper names and indexicals in relation to «intesionalism»). What has happened seems rather to be a change of methodological perspective under the threat of communication-theoretic scepticism prompted by underdetermination-problems. Thus the various attempts to articulate a theory of «direct» (but not immediate, as one should always add to avoid facile misunderstandings and dismissals and the fast search for refuge in some kind of «causal connection between sign and world» as an answer to the question: «but how the devil does a word get a grip on a thing»[14]) reference differentiate not only between «genuine names» and «definite descriptions» (as for the singular expressions) but also (as for the general expressions) between «natural kind terms» and «usual general terms» (or «n-criterion words» where n is the number of criteria you consider to be sufficient to determine the reference[15]) and (as for the intentions

[13] In «Essentialism and Reference», in: Schilpp, P./ Hahn, L. (eds.): *The Philosophy of W.V. Quine*, LaSalle ^{3}1988, pp.97-113.

[14] For this kind of shortcut see Devitt «Against Direct Reference».

[15] This is a liberal allusion to Putnam's term «one-criterion-words» (cf. «Is Semantics possible?», in: *Philosophical Papers 2. Mind, Language and Reality*, CambridgeMA 1975, pp. 139-152) as to denote the class of general terms having necessary and sufficient conditions for their application or are

for communication) between the «referential» and the «attributive» use of signs. It seems to me that some kind of these distinctions will appear as soon as you try to find out what it is that makes expressions of a determinate grammatical kind behave in a certain way (e.g. are counterfactual-supporting, extendible, have an «open texture» etc.). In the following I will try to trace some pragmatic aspects incorporated in the reflections on «natural kind terms» as opposed to «n-criterion words», like we could call general terms which are supposed to be referentially interpreted by complete necessary and sufficient conditions for their application (whose extension *is*, in other words, once and for all determined by their «intension», which might best be understood as «semantical» or «interpretative rules», i.e. terms whose «meaning» changes or gets lost when you change the conditions).

To get an impression of a territory the best thing is to have a look at its inhabitants. So the question is: What terms can be or are classified usually as «natural kind terms»? And the second question is: what is it that they share to be classified thus? or: What more general differentiation does this distinction aim to reflect?

To the first question: in the various writings investigating «natural kind terms» the most common examples given are concepts of lower biological taxonomy (as «tiger», «cat», «whale»), certain operationally defined magnitudes (i.e. relations, like «meter») fundamental concepts and magnitudes of physical theory («theoretical terms» like «electron», «atom», «impulse») and everyday-language expressions for substances («water», «gold»). Contrasting to that usually it is pointed out that the following do not satisfy the conditions to be «natural kind terms» (i.e., are «non-rigid» or «disguisedly descriptive» or, as I proposed, «n-criterion-words»): conventionally determined family- and property-relations and concepts («father of», «bachelor», «owner of»), concepts definable by contrast to some contingently preexisting classification («vixen» as «female fox» but not: «fox» as «male vixen», if you take the classification as grounded in «fox»), concepts for mathematical relations («square root of», «third derived from»), concepts of higher taxonomical order («mammal», «vertebrate being», «fish») and complex descriptions of chemical substances («H_2O», although this is not always entirely clear, some theoreticians seem to assume implicitly that these are «rigid descriptions», i.e. substance *names* instead of descriptions of chemical theory).

If there should be any order in this (hopelessly incomplete) list, at least it seems to me that it is far from evident. Even rough distinctions like «concepts for relations and states in the social world» vs. «concepts for relations and states in the objective world» or «concepts for more or less observable entities» vs. «concepts for more or less unobservable entities» are only good for a first try to give the extension of natural kind terms. You could add without hesitation disposition predicates, which are doubtlessly not only present in the discourse about the «objective» world (as opposed to the «social» one) and others, for their logic and problems apparently do not differ too much from the supposed logic of natural kind terms[16].

«defined terms» in the strict sense of «definition» mentioned above in the text.

[16] Cf. Goodman, Nelson: *Fact, Fiction, and Forecast*, Hassocks ^{3}1979 [in the following FFF], p.45, fn.9. One could even adventure (see note 1) the

This might seem to ignore all the things that have been said so far in the theory of reference with respect to «underlying structure» etc. What I am alluding to is the alleged theorem of Kripke-like approaches that for a term to be a «natural kind term» there has to be some strong ontological commitment to some not yet specified entity or mechanism or structure which is shared by all individuals falling under the respective predicate (say, *X*). What this is supposed to mean is that there *is* an identity-relation between all of them which *would* make the terms «normal» if only there were a possibility to be *sure* once and for all, *what it is to be an X*. As this is *ex hypothesi* not the case, we have to commit us to its existence, even if unknowable. As further these terms are supposed to be counterfactual-supporting, the phrase expressing this commitment would have to be necessarily true if true at all.

I only want to indicate here some reservations I have that keep me from integrating this without modification in a description of the «direct (but not immediate) reference account of natural kind terms» as such. Apart from the (important) question if this is rather a surprising consequence of unproblematic assumptions about the behaviour of general terms in possible world discourse or an independent axiom with pending plausibility (as I, following Salmon[17] and Deutsch[18]

hypothesis that a treatment valid for «natural kind words» should be expected to be valid for dispositional predicates as well: both types of expression are supposed to be counterfactual-supporting and -demanding: to explain the application of a dispositional predicate you have to invoke sooner or later a counterfactual condition, which is structurally the same when you demand that a kind-word refer «to the same things in all possible worlds». Both can only be introduced by reference to a part of their supposed total extension and have defeasible application-conditions, i.e. are supposed to function even when not associated with an exhaustive ncessary and sufficient condition for application. The best explanation of their use, i.e. to determine whether a given individual is or is not a such-and-such/has or has not such-and-such disposition is in both cases intimately tied to the best theoretical account available (this has been argued by W.K. Essler and R.Trapp in «Some Ways of Operationally Introducing Dispositional Predicates with Regard to Scientific and Ordinary Practice», *Synthese* 34 (1977), 371-96 and by Essler in «Some Remarks Concerning Partial Definitions in Empirical Sciences», *Pacific Philosophical Quarterly* 61 (1980), 455-62). I leave matters as confused and provisional as this because a thorough examination would demand its own place. However, see fn24 for some more details and section VI. for some speculations.

[17] *Reference and Essence*, Princeton 1981, Appendix II.

[18] «Semantics for Natural Kind Terms», *Canadian Journal of Philosophy* 23/3 (1993), p.404/405, where he shows that in a proper model-theoretic analysis of natural kind terms (his system NK) «the rule of *necessitation* [that is: $\phi \models \Box\phi$, A.M.] fails» (405). The important consequence this has for the usefulness of an «orthodox» reference-theoretic account (i.e., one making essential use of the notion of «rigidity» to model the behaviour of natural kind terms) of empirical classifications he stresses in «Semantical Analysis of Natural Kind Terms» (in: *Topoi* 13, 25-30) where he concludes: «It seems to

am inclined to think), this transition from truth to necessary truth seems to betray in a certain way the initial intention of such theorizing: namely to give an account how terms refer which do not at all, that is: *neither* factually *nor* counterfactually, have a necessary and sufficient condition for their application, i.e.: are simply underdetermined. To answer to this important question: well, they are determined, we just do not actually know what it is that (causally or however) does this, but suppose this thing, seems to eschew the question instead of answering it. There is undoubtedly something right in this answer, namely, that underdetermination confronts us with the unavoidability to reflect on what we suppose when going on to use the terms *as if* they were *totally* determined when we have determined their reference *somehow*. What seems wrong about the specific answer is the assumption that there has to *exist something which makes them determined terms independently of our decision to treat them as such*. Reflecting on what we do when we use terms as described in reference-theory and what it commits us to does not, from the outset, have necessarily to result in some outright ontological answer. Rather it would seem to me that this would be a surprise. What is to be expected by this kind of investigation is, in my opinion, not so much information about what the world is really like as what status is that we have to give the world as to be able to understand what we do when we are «simply going on to refer to the same with changing criteria of identifying it as such». To put up counterfactually some «ultimate identification» that legitimizes our doing so by telling us: «if some sentence like «a=b» is true, then it is necessarily true and thus this will be how the world is like with respect to a’s» does not really solve the question of how we succeed to keep track to one another and most of the referents *before* or *without* that substantial knowledge. As to confuse the case a little more: there certainly *are* a priori conditions that do permit us to do so, but they are, as I hope to make clear in the following, more general or *formal* and less demanding at the same time.

If there has to be drawn, then, some distinction between two ways to use general terms that corresponds to the distinction between the two uses of singular terms, it has to be looked for in metasemantic restrictions to the effect of distinguishing admissible and unadmissible interpretations such that in the case of admissible interpretations referential transparence and extensional determination by necessary and sufficient conditions do not coincide (i.e. where there is, for every model in a correct interpretation for a term *G(x)* some model for every necessary and sufficient condition *A(x)* for its application such that *G(x)* is satisfied by different individuals than *A(x)* in that model, that is, where the sentence «For all x: G(x) ↔ A(x)» is false). Thus these terms would qualify as special for being underdetermined in the sense that there is no criterially interpretable (or even «analytical») description of their extension which remains under all circumstances coextensional with the extension intended by the application of the term[19]. Such terms then *do not logically*

me that the semantical concepts of rigidity and nondescriptionality are secondary to that of an important property.» (p.30)

[19] The similarity that can be sensed here between the so-called «model-theoretic argument» given by Putnam and the conditions that give rise to the theory of direct reference is, in my opinion, not casual. It shows that Putnam’s argument, as given in *Reason, Truth, and History*, ch.2 and the proof in the appendix to the book, and its various variants, far from making him a

imply any determinate description of the extension for all applications, as judged by empirical adequacy and/or communicative success (although there might be for every application-instance one «contextually correct» description of the extension — but this, and this is the point, is not to be seen as an imprecision as to the relevant part of «meaning», as reference is to remain intact).

How do we fix the reference of such a term and how do we come to the assumption that it is referentially transparent? According to most theorists this is done by some sort of «baptism» or «dubbing» in the following way: you take some representative sample of the reference (in the case of singular case this question is simple, because there is only one individual, thus only one representative sample) which is a subset of the set of individuals falling under the term (say «tiger») and introduce the term by some remark to the effect: «this is a tiger, that is a tiger» and so on. Furtheron the term is (in the vocabulary of the person who has been taught the term or in the vocabulary of the language to which it has been added) supposed to refer to all individuals «like the ones in the sample». It keeps its reference intact either by continuous, unramified use (historical chains) or by thus getting glued to some causal mechanism which consists in something like «If tigers exist at all: whenever there is a tiger or meant a tiger and the word «tiger» employed, then there is a tiger (respectively: some organism with the genetical structure such-and-such) referred to» or by both. But baptism and causal chains are not the only possible interpretations of the pragmatics of successful reference fixing and keeping. Putnam also admits operational specifications (thus theoretical terms get covered as well) and in principle nothing seems to prevent any successful way to fix the reference to be legitimate: as the aim is only to specify something out of a set as paradigmatic, every means, linguistic or not, contextual, theoretical or whatever that accomplishes this, seems sufficient. This also seems to be implied by the fundamental fact wherefrom an alternative theory of reference gets its inspiration: if there is no one description that guarantees the reference a priori, then every one of them that fixes it in fact is correct, and as it does not depend on any description, even non-descriptions (in the given language) can be so. It is simply an empirical question how reference fixing is accomplished in fact, and baptism is just one model of a possible solution for the case of the introduction of a new term into the vocabulary of a given language (or idiolect). The same is true, it seems, of the «contact with the reference» that an individual is supposed to have as to get enabled to apply the term correctly. This can be helpful in the case of some sort of objects, namely the ones which can be perceived directly (or at least, «directly» relative to the language into which the term is to be introduced), but need not be literally the case in general. What is important is that the reference gets sufficient specification in the context of the introduction as to enable a speaker not to confuse cases of future application; and

«renegate» to realism (as M.Devitt would have it in «Realism and the Renegate Putnam», in *Nous* 17 (1983), pp.291-301) or committing him to transcendent idealism, shows (assuming that his reference theory is the core of his realist point of view) how realism is demanded for by paradoxical conditions *within* our practices when they are described in the traditional, semanticist way: the need for a new approach to reference is *prompted* rather than *risked* by the model-theoretic argument, it seems to me.

this can be accomplished, according to the case in question, in various ways which need not necessarily demand the presence of an individual of the extension[20].

Be that as it may, after a successful introduction a term is «referentially transparent» in the sense that we are, as all users of the term in question, supposed know that *there is a kind of things* that have (according to the best of our knowledge) some common trait and *to every individual of which one refers with the term*, e.g. all tigers. The set of all tigers, however, is not determined by any of the descriptions at our disposition that made us familiar with some of them, and therefore this type of fixing of use is *no consequence of the specific determination of the extension* accomplished by some description. We know, to put it a bit differently, that under different correct determinations of the extension under different circumstances the set of individuals falling under the term might differ, but we suppose that every individual in each of this sets has to be a tiger. As the number and structure of all possible determinations is, in view of the future and alternative states of the world, indeterminate, an effective way to give *the* extension is normally not to be expected. We could call this the *descriptive inexhaustability* of natural kind terms[21]. And it is exactly this information of the descriptive inexhaustability which is essentially part of our knowledge of the «meaning» of such terms[22]. It consists in our expectation that their reference is not covariant with «possible worlds», that is: alternative descriptions of the world in which there are individuals of this kind. In that sense we could explain this as a *normative* trait of the use of such words in the following way: we keep the interpretations of these terms constant through changes when we employ them, i.e. consider them to be referentially transparent, even though we do not (and often cannot) expect to be able to indicate the total extension, the product of the extensions under all circumstances (e.g. by some universal criterion of application), i.e. even though we assume their extensional opacity (relative to the possibilities of our language). The latter implies directly that there is no (semantical or other) fact that can be held uniquely responsible for the justifiedness of our referential expectations and presuppositions: these concepts do have, from the point of view of their *use* a regulated and concerning their reference constant application, but this invariance is (in general) not founded in any invariant condition of constatation of pertinence or characterization of the members of the extension.

[20] The decisive steps to answer these questions would be some account of the representativity of a sample as much as a general account of what it is and how we know or suppose that some specification is sufficiently exact *in the introductory situation*. But this is far too complicated to be treated in this article.

[21] Thus the alleged «non-descriptivity» of natural kind terms would not be, as is often suggested, a result of some capacity of language to refer without any descriptive context but rather one of the continuous possibility of revision and conceptual change: there are not too few, but too many possible descriptions of the extension as to guarantee by this criterion referential transparency.

[22] This has been argued by Goosens and later by Deutsch (see below).

Kind words as characterized up to now thus seem to be unseparably linked to knowledge-changig practices, for the *central rule for their use* would then be to know exactly this: that they, although introduced and explained by descriptions, are not equivalent with them. The «original concept» which gets introduced in some vocabulary *together with the implicit or explicit information that it is a kind term* is almost empty. In that sense H. Deutsch[23] remarks: «It does not take much to be *that kind of thing*. (...) if we were armed with only the original concept of cat [his example of a kind concept, A.M.], we wouldn't know much about cats. (...) The possibility that cats are really automata is rooted, not in our ignorance, or possible ignorance, of the nature of cats, but in the meaning of the word 'cat' — the original concept of cat.» (p.409)

The problem with the talk about «reference» in connection with that type of general terms is obviously, as «reference» and «satisfaction of a description» do not coincide here *ex hypothesi*, to characterize what it exactly is whose existence is supposed to be able to refer to it. This has always been the decisive question where essentialism lurks, which might be no problem for philosophers who believe in real essences and try to prove their existence by some theoretical construct or other, but it certainly is not uncontroversial. How does it come about?

An important premise for the explanation of reference in this manner seems to consist in the idea that, given that a sufficiently well introduced term is to be considered as part of the background knowledge in a certain situation, one has to suppose the *description-irrelative (i.e. independent) existence* of some «object» (or better: reason) of referential transparence, which has to be the result of some (generally *unknown* and often *supposed to be unknown* and therefore *not completely statable*) general trait common to all individuals that are members of the kind in the case of general terms (as to be able to refer with the general term to each of the individuals that are memebers of the kind). W.K. Goosens[24] dubbed this

[23] «Semantics for Natural Kind Terms», in: *Canadian Journal of Philosophy* 23/3 (1993), pp.389-412.

[24] «Underlying trait terms», in: Schwartz, S.P. (ed.): *Naming, Necessity, and Natural Kinds*, Ithaca/London 1977, pp.133-54. Quine uses a similar term in connection with his explanation of the functioning and purpose of dispositional predicates («Necessary Truth», in: *The Ways of Paradox*, Cambridge, MA, ☺1975, 68-76) and clarifies their close resemblance with natural kind terms in «Natural Kinds» (in: *Ontological Relativity and other essays*, NY 1969, 114-38)). It would be worthwile investigating further Quine's conceptions and to compare them with what has been said in natural kind term reference theory. This is so because, following Quine's arguments one can see without difficulty a parallelism between dispositional predicates and kind terms and the evident importance of both in scientific practice, i.e. their epistemological import. Some indications may suffice to justify this claim: Quine calls (in «Necessary Truth») the counterfactual conditional-discourse underlying the use of dispositional predicates as indispensable for imputing dispositions on a domain and, above all, for the innerscientific practices of prediction and formulation (and interpretation) of hypotheses (p.73, 69), and describes its general epistemological structure as follows: «In general, when we say 'If x were treated thus and so, it would do such and such', we are

characteristic of kind terms as the presupposition of some «underlying trait». The problem was, that possible world-invariant properties or traits seemed to be directly identifiable with «essential properties», i.e. attributions of necessary truth. This is due to the attempt to draw for the distinction between natural kind words and n-criterion words on the distincton between possible world covariant and non-covariant properties. And this, if interpreted *ontologically*, leads fast and neat to talk about essences. Thus it seemed that natural kind terms might be «non-descriptive» concerning contingent properties, but surely had to be descriptive concerning «metaphysically necessary» properties. What stands in question is not, of course, the logical correctness of this conclusion when you accept some sort of Kripkean interpretation of modal logic. What is questionable is from where you want to apply it: if it is applied or interpreted in any absolute sense, then you get to essentialism. But when this conclusion gets situated within the description of the rules underlying our discourse in hypothesis-accepting practices as their interpretation-theoretic structure, then «necessity» and «essences», once gotten in the scope of reflexion, get

attributing to x some theoretical explanatory trait or cluster of traits.» (ibid., my italics). This attribution has the following status respectively function within a given corpus of knowledge: «the [disposition-, A.M.] term has been a *promissory note* which one might hope eventually to redeem in terms of an explicit account of the working mechanism.» (p.72, my italics) This suggests that the hypothesis to the effect of some «working mechanism» or a «sub-microscopic structure» (ibid.) in the case of chemistry, in general of an «explanatory trait» (ibid.) is less to be interpreted as a serious hypothesis about the furniture of the world in itself than as a provisory, hypothetical and confirmable ontological posit with pending justification: «In the necessity constructions that impute dispositions, the generality lies along some *known or posited explanatory trait*. (...) They turn, still, on generality. But they *turn on theory, too*, precisely because they fix upon explanatory traits for their domains of generality.» (74) The acceptance and the concrete content and structure of these «promissory notes of common traits» is thus shown to vary from epoch to epoch, depending on the accessible «underlying theories» about what is possibly to be counted as a component of such a «common trait»: «What kind of account of a mechanism might pass as explanatory depends somewhat, of course, upon the general situation in science.» (72) This means that the use of dispositional classifications is comparably weak and relative to other, more fundamental or at least
already approved and accepted classifications which in turn are seen to determine ontology. This is so because until there is no lawlike statement (or, according to the discipline) something functionally equivalent to it, the assumption of some such «ontological hypothesis» is nothing more and nothing less than a hypothesis with uncertain justification. Now, such general statements of law are known not to be inductively confirmed in any direct way (cf. e.g. W.K. Essler: *Induktive Logik*, Freiburg 1970, chap. V.4.); they are thus best be seen as belonging to the (contextual) *a priori* part of a theory as a whole which is rather than a consequence, a precondition of the investigation in the structure and content of the world.

an acceptable and even explanatory reformulation (and thus loose, perhaps to the disconcern of some, much of their metaphysical robustness[25]).

Roughly the change is this: invariant properties can be seen as «important» properties, where «importance» depends on the explanatory aims of the respective practices. The relation of some pretheoretical notion of «essence» and «importance» can then be seen in the following way: «important» properties may not coincide with the «real essences», because we might restrict our interest to the cases where some property invariably occurs (e.g. to investigate its connexions with some other), i.e. voluntarily restrict the given domain of things of a kind to things of a kind in events of a type. Then this property would be «important» but not «essential» to things of the kind in general. On the other hand one should expect that «essential» properties (whatever they are supposed to be) should count always as important, when known. And this is the point: it is a commonplace that «essentiality» is something we cannot «get to know» by any standard scientific investigation. Thus, seen from an epistemological angle, the intersection of the class of essential properties and the class of all knowable or investigable properties is the empty class. The most natural reaction to this is to put up some principle like: it should be the case that the properties considered by us as «important» for the description of a kind be its «essential» properties in the sense that our generalization concerning its members be true. This, in turn, can only be argued for inductively. Our belief in the truth of the generalization can only be *confirmed* with reference to a subset of all members of the kind and thus can also be *infirmed* and even be *falsified* with reference to it. Then it might be rational to drop this classification, even though at some later stage additional information or a new explanatory approach account for the reason of this infirmation and the classification can be «revived».

Nevertheless the conviction that members of a kind, if it actually is a kind, must have some *absolute* common trait, has a place in these practices. But it does not follow *logically* alone from some given postulate (that is: *that* is not the interesting point) but is itself a *counterfactual* statement with normative content about the «grammar» or functioning of *all kind terms in general*. Whenever we have reason to suppose of some concept that it is a kind term, then this means that we know that some hypothesis to the effect that there is some trait common to all its members (however they be identified) — some «homogenity» in the domain — is

[25] This is stressed by Putnam in «Possibility and Necessity» (in: Putnam, H.: *Philosophical Papers 3. Realism and Reason*, pp.46-69) where he remarks: «the 'essence' that science discovers is better thought of as a sort of *paradigm* that other applications of the concept (...) must *resemble* than as a necessary and sufficient condition good in all possible worlds. This should have been apparent already from Quine's criticism of the analytic-synthetic distinction.» (p.64) That is: if you want to design a theory against the underdetermination-problems stemming from this criticism, this theory should not imply theorems to the effect of reproducing the very target of this criticism. So goes only half the way when he affirms in the same context: «saying that 'Water is H_2O', or any such sentece, is 'true in all possible worlds' seems an *oversimplification*» (p.63); it is simply *just as inadequate* as saying that some such sentence is 'analytic' and *subject to the same criticism*. It is, in other words, if theorem of a theory, part of an *inoperative* theory. In case of being an axiom, one should consequently look for ways to avoid it.

valid in the language where this concept is used. We are confronted, then, with a restriction concerning admissible interpretations (and not in the least with «limits of our world») as asked for for *supposedly empirical reasons*: as long as we assume the adequacy of a given classification, of which some kind term is part, only such «worlds» are admissible domains, in which we refer with the term only to individuals which are actually members of the kind and to all of them in the respective world, according to our best theory, i.e. where this homogenity claim is satisfied.

Thus substitutivity of all tokens of a kind word in all contexts is, as in the case of «genuine names», no logical consequence of the determination of application for this general term, but a counterfactual («grammatical», as Wittgenstein might remark) claim concerning te *functioning of kind words*, namely: that they *should* be substitutable in *all* contexts (including the modal ones) and that all ontologically relevant operations (identity, quantification etc.) are valid for them even in case that there is no *analytical* or *absolute a priori* definition for them.

There are, then, certain traits of practices that demand (or at least: whose participants regard) it as a constitutive fact of their possibility that the extension of some terms cannot be completely given or given by a mechanical procedure alone but nevertheless there is a provision for their empirically clear use. Dagfinn Føllesdal[26] has provided, for the case of «genuine names», a list of conditions which prompt this type of interpretation-theoretic entities: «Names are normally introduced for the following three purposes:

(i) When we are interested in *further features* of the object beyond those that were mentioned in the description that was used to draw our attention to the object.

(ii) When we want to follow the object through *changes*.

(iii) When we are aware that some or many of our beliefs concerning the object are *wrong* and we want to correct them.» (S.108)

Contexts of use like the ones described by Føllesdal and the modalities hinted at before could be called, following Goodman, contexts in which we want to «project» predicates and the statements formed with their help. Those are contexts in which the «proceeding from a given set of cases to a wider set» (FFF, p.58) (where «set» can be understood as set of applications) not only is being made in fact but is furthermore part of the *normative expectations* imputed on a competent participant and is seen as (at least retrospectively) *rationally justifiable* and a legitimate proceeding is interpreted as a *learning process*. The paradigmatic cases in question are undoubtedly contexts in which one uses *inductive* procedures. In that sense, the foregoing could be seen as the attempt to describe the projectibility-conditions for the case of natural kind terms, which constitute an especially interesting case because what is intended with the term «natural kind term» seem to be tha classifications in use that lay on the ground of the practice of natural science.

3.— Goodman’s paradox

[26] Cf. «Essentialism and Reference», in: Hahn, E./Schilpp, P.A. (eds.): *The Philosophy of W.V. Quine*, LaSalle ^{3}1988, pp.97-113.

Goodman's paradox is usually situated, as its inventor did, in the context of questions concerning the justification of inductive reasoning and, more specifically, of confirmation theory.

As I only want to draw the attention to some points where I think the problems that gave rise to reflexions about a new approach in the theory of reference and the problem discovered by Goodman[27] coincide (or at least converge to the same reason), I will not suppose anything very original under the term «induction». When in the following there appear expressions like «inductive practice» or cognates, this is to be understood as a practice guided by some canonical method to relate in a systematic form singular experiences with generalizations and expectations which do not follow deductively from those. Roughly, such a method will permit to consider it to be rational accept some hypothesis or sentence as true if there is a sufficient number of positive instances at disposition. Such a step from a sum of singular true experience-describing sentences in the position of premises to some other sentence held (to degree n, if you like) true[28] is then an «inductively valid inference» and the hypothesis is to be seen as «confirmed» (to degree n, if you like) by the experience at hand. Among these one can decide the two groups of

a) singular predictions of the form

(a) $a_1 \in F, a_2 \in F, ..., a_n \in F \rightarrow a_{n+1} \in F$ and

b) inferences from singular data to general hypothesis of the form

(b) $a_1 \in F, a_2 \in F, ..., a_n \in F \rightarrow \forall x(x \in F)$

[27] This idea goes back to Quine's article «Natural Kinds» (in: *Ontological Relativity and other Essays*, N.Y. 1969, pp.112-38), where he treats dispositional terms, kind terms, counterfactual idiom, similarity grades and simplicity as a problem-cluster, for which he suggests that a clarification of one of the problems would have immediate consequences for the treatment of the others. However, I have the impression that Quine sees this problem-cluster as a sort of residual sphere of «second order» intensional talk which will be superseded as extensionalist approaches get better. This does not seem to be the case, for the approaches of Føllesdal and Kripke to some of the named problems does not make use of intensions in any suspicious way; quite on the contrary, Kripke's model-theoretic semantics of the modalities converts the whole idiom in a perfectly extensionalist language. And it was exactly there where the necessity for a distinction between kind terms and general terms arose. So it seems that the problem remains under extensionalist treatment.

[28] This is, of course, not to say that there are grades of truth. There are supposed to be grades of acceptability, measured by some measure-function (usually supposed to be some modification of a probability calculus), but not of truth, for what is to be accepted is the *statement* in question, i.e. true sentence. And this taking to be true of some determinate sentence is considered to be more or less rational, according to the output of the canonical method.

The singular statements on the left represent the data at hand consisting in the results of an investigation of the domain of individuals concerning the property *F* and the clause on the right is the hypothesis.

The question which interested Goodman was whether firstly every two coextensional descriptions of the experience at hand result under the same canon inevitably in identical grades of confirmation and secondly if, therefore, there is something like «objective» learning from experience which functions without more prior knowledge of the concepts used to describe the experience than knowing that they exhaust all data.

Against these two ideas — the idea of an absolute measure function and the idea of the independence of inductive inference from linguistic descriptive means — Goodman construed in «A Query on Confirmation»[29] the following example:

Take a bowl full of emeralds. Until some moment t there have been drawn 99 green balls from the bowl. What would be, pretheoretically speaking, the correct singular prediction about ball 100? According to scheme (a) we would infer (correctly):

$a_1 \in \text{green} \wedge a_2 \in \text{green} \wedge \ldots. \wedge a_{99} \in \text{green} \rightarrow a_{100} \in \text{green}$.

That is: «'$a_{100} \in$ green' is true» would be more probable or better confirmed by the available data than «'$a_{100} \in$ green' is false». Now Goodman construes the following predicate «grue» which is, concerning the available data, coextensional with «green». The definition is:

DGRUE «$\forall x(x \in \text{grue}) \leftrightarrow [(x \in \text{green} \wedge x \in \text{drawn until t}) \vee (x \in \text{blue} \wedge x \in \text{drawn after t})]$».

According to rule (a) one should expect that «'$a_{100} \in$ grue' is true» would be more probable or better confirmed by the available data than «'$a_{100} \in$ grue' is false». This implies, of course no logical contradiction, taking in account the definition given: why should a_{100} not be grue if the premises are $a_1 \in \text{grue} \wedge a_2 \in \text{grue} \wedge \ldots. \wedge a_{99} \in \text{grue}$? A contradiction raises only if we try to infer *simultaneously* both ways, i.e. try to apply «green» and «grue» simultaneously to a_{100}. Nevertheless it is clear that our previous experience with colour-terms and their application conditions do not contain anything like the expectation of sudden changes because of the lapse of time (without changing something else within this lapse of time, like switching on a redlight bulb) and our experience with precious stones does not admit of too much variation in colour without a variation of the sort of stone and our conception of regular drawing-bowls dictates that there be no variation in the data just because of the lapse of time. True as all this may be, it is not sufficient to rule «grue» out as badly defined over «green» as correct and the «grue»-hypothesis over the «green»-hypothesis. The only additional condition you need for the definition of the new predicate to pose a problem to «normal» predicates is the quite

[29] In: Goodman, Nelson: *Problems and Projects* [in the following PP], Indianapolis 1972, pp.363-6. A precision of this argument resulting from the subsequent discussion with Carnap about this article can be found in «On Infirmities of Confirmation Theory» (PP, pp.367-70). The most famous version of the problem is probably the one in FFF, chapter 3 («The New Riddle of Induction»).

modest premise that there are more objects to be examined than the ones contained in the available data[30].

The problem is general and not artificial, because there is a general rule[31] for the generation of such predicates which does not infringe any accepted rules of reasoning straightforwardly. It is simply a problem about a distinction between predicates apt or not apt for inductive inferences which permit learning from experience. This means that a consideration of the premises that lead to it (and a fortiori, a search for premises that prevent it) will show premises implicitly invoked in the cases that we pretheoretically consider to be satisfactorily solved.

In the following I will not discuss all the numerous proposals for a solution, dissolution or *a priori* rejection of Goodman's «new riddle of induction»[32]. Especially I will not discuss Goodman's own approach to a solution in form of a theory of «entrenchment», because it draws only on *custom and origin* of the predicates. These are, however, facts about behaviour, describible in predicates which, in turn, can be submitted to a «corruption» by the general rule: we cannot know whether «projected» is corrupt or not without some further information about what it is that justifies projection of projectible predicates and unjustifies projection of corrupted ones. The problem is about *validity in general* and not about *the empirical fact of projection of a particular historical period* and can thus — *pace*

[30] Cf. Putnam, H.: *Representation and Reality*, Cambridge MA, 1988, p.13 where he remarks on the occasion of interpreting the changes in the specification of the concept 'electron' within a «story of successive changes of beliefs about the same objects» (namely Bohr's various descriptions of them): «to treat all (...) occurences of 'electron' [within this process, A.M.] as synonymous as is involved in his [Bohr's, A.M.] decision to treat later research programs in the story as *extensions of the earlier ones* (...) plays a central role in theory evaluation. In fact, treating 'electron' as preserving at least its *reference* intact through all this theory change and treating Bohr's 1934 as a genuine *successor* to his 1900 theory is virtually the same decision».

[31] This point is, to my knowledge, due to W.Lenzen (*Theorien der Bestätigung wissenschaftlicher Hypothesen*, Stuttgart 1974, p.174ff., esp.183, fn5) That is to say, the new predicates construed by a definition like the one for «grue» («corrupted» we might call them, following W.K. Essler's terminology in «Corrupted Predicates and Empiricism», in: *Erkenntnis* 12 (1978), pp.181-7) do trivially coincide with the «normal» ones in case that the second clause (after the 'or') in DGRUE is false because of a factual truth like

[32] This rule is the following (cf. F.v.Kutschera: «Goodman on Induction», in. *Erkenntnis* 12 (1978), pp.189-207): Take the two principles induction (a) and (b) and any predicate *F* and any set *A* of objects such that $A=\{a_1, ..., a_n\}$ is the set e.g. of all objects tested for F until t (or, more general, a non-empty real subset of all objects in the supposed universe such that there is an $a_i \notin A$ with $i \neq 1, ..., n$), that is, the available data. Then, this is Goodman's argument, there is a predicate *F'* such that $F'(a_j) \leftrightarrow F(a_j)$ for $j=1, ..., n$ and $F'(a_k) \leftrightarrow \neg F(a_k)$ for all $k \neq 1, ..., n$. Formally this is $F'x :=_{df} x \in A \wedge Fx \vee \neg x \in A \wedge \neg Fx \vee$. The contradiction arises for a simultaneous application of (a) or (b) for a_{n+1} with respect to *F* and *F'* which seems justified, for $Fx=F'x$ for *A*, which gives $F'a_{n+1}=\neg Fa_{n+1}$ and Fa_{n+1}.

alleged virtuous circles — not be on a way to a solution if this way is absolutely a-normative[33]. Thus Goodman is in my opinion completely right when he writes: «Any argument that the initial choices of projectible predicates are determined by some non-random operation (...) requires showing that these *predicates* are distinguished by some common and independent characteristic (...) that can be correlated with such an operation (...). The *unavailability* of such a characteristic (...) is just what gives rise to the riddle.» (PP, 358, my italics). There is just one condition which sounds a bit like Goodman's attempt but is in effect of totally different sort, namely the following adequacy condition for a solution: any definition or theory of «projectibility» has — in analogy to a theory of truth in semantics — to yield all predicates as projectible that have been judged pretheoretically as sucessfully projected[34]. Nevertheless the change of perspective from a purely syntactic-semantic treatment to a pragmatic one (as no condition as to the properties of the *signs themselves* without consideration of their *use* or *application* helps) that Goodman proposes in remarks like «entrenchment derives from the *use* of language» (FFF, p.95) seems absolutely right. This is the line I want to follow in the following a little bit. I should make clear from the beginning, though, that I consider the theory of entrenchment as one of the possible specifications for a strategy towards an elucidation of what the use of language in contexts where we distinguish valid from invalid inductive inferences, and furthermore as one which is (for the reasons given above) not too promising. Thus the following is not to be seen as a contribution to entrenchment theory, and I fear it will not even be sufficient to indicate a determinate strategy for avoiding Goodman's problem. What follows is rather intended to apport some more evidence to the suspicion that there are common points in the desiderata raised by reference-theoretic and projection-theoretic problems.

[33] To get an impression of the impact caused by it, may it suffice to recommend the excellent collection of essays on Goodman's paradox provided by D.Stalker: *Grue!* (Chicago/LaSalle 1994), especially the exhaustive annotated bibliography of texts in English on the problem contained in it.

[34] Goodman himself says that in view of this problem the aim has to be to reach a «dichotomy of predicates» (FFF, p.80). The insufficiency for an answer to this question of the resources given to us by past behaviour is stressed by him in his critique to Hume when he says: 'Hume overlooks the fact that some regularities do and others do not establish such habits; that predictions based on some regularities are *valid*, while predictions based on others are not. (...) To say that valid predictions are those based on past regularities, without being able to say *which* regularities, is thus quite pointless. Regularities *are where you find them*, and you can find them *anywhere*.' (FFF, S.82) This obviously applies *mutatis mutandis* to descriptions of «induction-*regularities*» found in our culture (be they or not reached by reflexion: the question in point is whether they have *normative* import or not). What is demanded is a *general procedure to distinguish two types of predicates* in the structure of which one could find some set of interpretation-theoretic presuppositions of valid inductive inferences (as opposed to invalid ones); this is obviously impossible if one is limited to *particular and contingent* descriptions of the set of all valid inductions.

4.— Aspects of the interpretative theory underlying projectible predicates: Some remarks on the conditions of the distinction between projectible and non-projectible predicates

The reason why I do not think that my observations in this part will constitute a solution (or even only the nucleus of such) to Goodman's paradox is precisely that I am inclined to think that there is no outright solution from an absolute vantage point. *Absolutely* seen, the paradox is, due to the fact that it is generated by a perfectly admissible general rule, unavoidable for any empirically interpreted predicate. Predicates or classifications can only be justified *from within a practice* and the same predicates can always, *in a reflexive metalevel* of the language used in this practice, be «corrupted». It would not even be correct to exclude the rule leading to the corruption and the premises necessary for its derivation *a priori* as bad in general: there are corrupted domains where *only* corrupted predicates (and their respective confirmation-methods) are adequate, moreover there should not lurk any intractable problems if we *know* how the corruption has taken place: in that case we could e.g. simply modify our definition and make it conditional on this mechanism (which would be a kind of errorlevel-fixing)[35]

So what seems to be called for is, in my opinion, a view of how this activity of «corruption of predicates» is restricted, when necesary, and what the world and the language is supposed to be like to do this justifiedly. These presuppositions are, thus, presuppositions of inductive rationality. I will try, then, rather than to solve, to *reconstruct* Goodman's riddle in such a manner that we get a glimpse of why we do not always have to struggle with the paradoxical consequences it can have and what the ontological and epistemic commitments might be like that we have to make to do so.

Now, what does it mean to change the perspective towards a point of view that throws some light on the *pragmatic* conditions of interpretative procedures in inductive practices that help us understand the process of learning from experience as a rational one? First, it means indicating oversimplifications: «The fact is that whenever we set about determining the validity of a given projection from a given base, we have and use a good deal of other relevant knowledge. *I am not speaking of additional evidence statements*, but rather of the record of past predictions actually made and their outcome.» (FFF, p.85)

For the reasons given in the last part, Goodman's last remark is not very convincing as it stands. A possible treatment of the difficulties raised by the problem cannot be expected to work if it is limited to the invoking of the *fact* of background knowledge, and *mutatis mutandis* neither by any structured description of some background knowledge. It has to start not only from *use* but from a representation of the *rules for use* of the predicates, to indicate further some rule contained in all sets of rules for projectible but not contained in the ones for the unprojectible predicates (or something like that). For this to be efficient it is, again, insufficient to put up some classification of *de facto* successful words and their understanding (or «meaning»). The aim has to be to indicate the (ontological, cognitive and epistemological) structures implied by these rules that *explain success*.

[35] Cf. Hacking, I.: «On Kripke's and Goodman's Uses of 'Grue'», in: *Philosophy* 63 (1993), 269-95.

Having put things like this, the question seems to be: what are the *normative conditions invoked by and taken for granted in the application of predicates such that they do not prevent learning from experience*? These conditions would be, in turn, part of the background knowledge, but *not* (as Goodman correctly indicated) some substantial knowledge about the content of the domains. They would be of a much more general sort, like the «knowledge» that there is some domain and that there is a language and that both have to be related to each other in an interpretation that assigns objects to signs (constitutes reference) etc, and that the *intended domain* for application of a certain predicate is of a certain *structure*. They would constitute something like a cognitive matrix that is inevitable for the correct mastering of predicates in certain practices. We have already seen that on the purely syntactic or semantic level there is no difference between «grue» and «green»: you can define «green» in a vocabulary of «grue» plus an individual constant and vice versa. Both are *definitorily symmetric* and furthermore they are, with respect to the given data (in *A*) *eliminable*, i.e. in the description of all individuals in the data you can always substitute the definiens of the respective describing predicate *salva veritate*[36]. Both predicates become *asymmetric* in the case of a_{n+1}: a_{n+1} is grue iff not green and vice versa. Now, this could be described as a process of becoming extensionally opaque although there would be, *with respect to the given data* the possibility of an identical necessary and sufficient conditon for use. If we take *A*, as plausible, as the set of our possible paradigms for introductions of the predicate, then there is always a manner to go on with the application of the predicate in question which resembles «grue». *On logical grounds* there is no means to prevent it. It seems, then, that we are confronted with the same phenomenon as in the second part. An indication of some

[36] A case where it might be reasonable to keep or construct a corrupted predicate could be, for example, the case of some «objective» change that would, however, for its particularity, not call for a change of theory, but nevertheles for a modification in the homogeneity-supposition. Imagine for example human population of a certain specific genetic structure inhabiting an area with active volcanoes. One day one of them erupts and this eruption causes testable changes in the genetic material of the children of the members of the population that survived the catastrophe. In that case it would be irrational to expect the predicate of genetic theory, say «to have characteristics F, G, and H in the genes», which was coextensional with «being a (geno-)typical inhabitant of area V» (and even theoretically more exact) before the eruption to be projectible afterwards. That means that it would be irrational, *knowing the «rule» of corruption*, to go on using the genetic predicate of before the eruption because it used to be perfectly projectible; rather one should take the «rule of corruption» in account and add it to the background knowledge (in the example this would be exactly a «Goodmanian» disjunction of the type: «someone is (geno-)typically V-ish iff he/she either has characteristics F, G, and H in the genes and was born before the eruption or has characteristics K, L, M and was born after the eruption»). One could even go on with this example and think of the possibility that this variation is not of dominant character and thus disappears, say, after the seventh generation so that the first predicate gets fully re-applyable (when all members of the variant-population have died). For the time between these two events it would, nevertheless, be inappriopiate.

necessary and sufficient condition (i.e.: identical examination procedures) based on *A* does not lead, as we go on applying the predicate, to the same results with the same objects: both predicates get «ramified», such that after t (or after a_n) the test-conditions to be satisfied to count as «green» or «grue» are mutually incompatible. If we assume that the examination method does not change and the objects after a_n give the same results as before, then all attributions of «grue» become false, i.e. «grue» has not been specified sufficiently in *A* to refer to grue things, whereas «green» has been sufficiently specified: it keeps its reference, whereas «grue» has still to gain a representative class to get its reference fixed: *A* was paradigmatic for the introduction of «green» but wasn't representative for the introduction of «grue». As both are interdefinable, this is always true for the respective definiendum and definiens: in the definiens there is an essential occurence of an individual which causes the non-representativity of the assumed sample, i.e. to be a sample for the predicate in the definiendum requires always *more* concrete information on the individuals of the *extended* domain and their properties. Now, our assumption that with «green» we can go on as before, i.e. that *every* token, independently if uttered before or after a_n, is substitutable salva veritate for every other, depends *a fortiori* on a different presumption (or «information»), namely that the universe is, with respect to «green», homogeneous enough that *A* can be seen as representative and thus serve as the class to introduce «green» in such a way that it becomes *definite* or *non-ambiguous*: it refers, if it refers in *A* to objects of a sort, then *always* to objects of *this* sort. «Grue», on the other hand, gets assigned as reference objects that have, relative to relevantly the same test-procedures (e.g. colour-analyses), *different* structure: there are some *F*'s and some non-*F*'s in its extension. To apply it with the help of these test-procedures, we have consequently not only to know the result of the test, but furthermore *which individual it is we apply the predicate to* to be able to determine the truth value of the respective sentence.

So we could say that relative to a remaining application-background, one of the predicate refers to individuals with relevantly the same structure whereas the other does not. If we assume now that Goodman is right in that «green» is projectible and «grue» is not, this can be expressed as saying that we assume for «green» that there is something which all green things have in common in a certain respect (e.g. colour), whereas there is nothing like that in the same respect within the same cardinality of a sample of grue things[37] (for they could have something different in common). If we suppose this of our universe of things, homogeneity in a certain respect of more abstract order within a system of classification (like: light green is a sort of green is a sort of colour is a sort of optical thing property is a sort of thing

[37] This follows from the definitions given above: If we take *A* as e.g. the class of all experiential sentences up to now and assume that there is no more individual than those contained in *A*, then *F*x=*F'*x (see fn30). This is quite obvious, because the individual expression needed to define either «grue» on the basis of «green» or «green» on the basis of «grue» does, in that case, not refer. Thus the condition «not drawn until t» is trivially true because there is no more thing to be drawn and the condition in the second part of the disjunction of the definiens is, because of the (in classical logic) trivial falseness of «drawn after t», also false. Therefore the definition of «grue» (i.e., in general, of *F'*) is satisfied, in this case, by all things that are green (i.e. *F*).

property is a sort of property), then «green» is referentially transparent relative to our colour theory whereas «grue» is not, for an examination of colour *alone* is not sufficient to determine whether a given thing is grue or not and thus our aplication of the predicate correct or not. Nevertheless we *do* have a necessary and sufficient condition for the application of «grue» on the basis of our colour theory and the harmless assumption of more things than the examined ones in our universe (which is, in turn, nothing more than a specification of the interpretation-possibilitating premise of a non-empty universe), viz. the definition given in the previous section.

What does all this amount to? It amounts to saying that having necessary and sufficient conditions for the application of a predicate does not mean to have projectible predicates. It is neither necessary nor sufficient for this. The presumption of the projectibility of a given predicate depends instead heavily on uniformity assumptions contained in underlying theories about the structure of things, i.e. assumptions about the world *understood as such from within a certain practice* (this is important, because our background theory could be corrupted and its background theory etc.[38]: *realism* in that sense is *always* an intrinsically internal presupposition for the *use* of predicates and, if hypostatized as *absolute*, immediately subject to corruption-counterexamples[39]). To put this point a bit differently: there is no description of any part of the extension (i.e. class of things to which a given predicate applies) which *guarantees continuity of reference* and thereby the projectibility of allegedly projectible predicates. Non-projectible predicates, on the other hand, are only applicable if we give a description of one determinate part of the extension.

The important thing is that nevertheless the projectibility or continuity of reference is *essential* for the possibility to learn of new things that they are like the known things and to learn, starting from their common features, about their differences, in short: to learn from experience. To make predicates projectible, we have thus to suppose a) that there are things that have the property we assign to them independently of our knowledge of them (that there are more than the ones we know, for otherwise there would be neither any interest for a property nor a possibility for the «grue»-predicates to come up, and thus no reason for scepticism about the referential efficiency of our predicates) and b) that there is something common to all things we apply the predicate correctly to. This is, however, a

[38] For example we could imagine (in the case of Goodman's 'grue') that there is only a finite number of additional things in our universe which does not exceed the number of green things -n; then the average homogeneity of the class of all grue things could never reach the average homogeneity of the class of green things because n does not become sufficiently little to be neglected. Otherwise the set of all grue things becomes (in the limit) almost indistinguishable from the simple (inductive) complement-class to green (*pace* the n green things in it, but if n is very much lesser than the total of all grue things in the limit, then there is an almost zero-possibility to get something green out of the class of grue things).

[39] This has been argued against certain attempts to dismiss Goodman's problem on grounds of «natural» categorization-systems which provide us with a conflict-blocking overhypothesis by J. Ullian in «More on 'Grue' and Grue», in: *The Philosophical Review* 70 (1961), pp. 386-9.

consideration from *within* a certain practice: «green» is, once introduced, referentially transparent and «grue» is not because we base our expectations on some available background theory which we do not question in this same moment: it is our «theory of how the world is» and contains some assumption to the effect that colour attributions are independent of time points and the selection of individual colour-porters as such: colours are «in the world» (of our not problematized background-classification), whereas the selection of individuals to be examined and its temporal structure is not (because it is never, according to virtually all theories of regular confirmation-methods). In that sense we could talk of a *realist* assumption of projection-practices, where the realism enters in form of the presupposition that there is something common to the things in general (the world) independently of how we introduced the term and afterwards introduce it: every representative class *within* our available data is such that it is *sufficient* to exemplify the property in question *for all* subsets of the application.

This is, however, obviously not a *substantial knowledge about the things in themselves* but a presupposition about the structure of the universe of discourse. And *this* is in turn not a presupposition linked to the predicates as such but to their *application* and *use* and the rationality assumptions related to them: if we want to learn from experience, we need projectible predicates, and if predicates are to be projectible, we have to presuppose the mentioned structure of the world. Without our interest in learning from experience, the interest for projectibility would have, in my opinion, no reason. This is so because complete ignorance (that is, the supposition that there are no more objects to be examined) blocks non-projectibility and complete knowledge (that is: including the ramification-conditions in the definitions of non-projectible predicates) makes it harmless. And the interest to learn from experience and its necessary conditions, that we neither know nothing nor everything about all things in the respect in question, named by some predicate, is, obviously, a presupposition of our rationality.

The knowledge needed for the use of projectible predicates thus cannot consist in a knowledge of *the* (factual or counterfactual) extension,, but only about the supposed common structure of all individuals that are members of the extension*s* in the plural that we successively determine under varying epistemic circumstances. This knowledge is, concerning every single predicate, a knowledge of the homogeneity to be supposed for the things in the universe in that respect, concerning *all* projectible predicates it is the «knowledge» *that* their domain has a non-corrupted (or known corrupted) homogeneity, in other words, that the predicate can be correctly applied to every individual object of it *independently of the fact of how and when it has been registered or identified*. The behaviour to use such a predicate *as* fixed, as *applicable to a determinate sort of things* is, then, to be seen as a rationality assummption concerning the use of a determinate sort of general terms.

It implies *epistemologically* that one has to suppose certain things *about the domain* as soon as one uses a predicate and presumes it to be projectible. This, in turn, means that the supposition is *independent* of the determination of the respective extensions *in* a certain epistemic situation, as long as the domain is supposed not to change. This *relative independence*[40] of determination of extension in an epistemic

[40] This is in fact the same as what Putnam says in his «model theoretic argument», in *Reason, Truth and History*, Cambridge MA 1981, ch.2. For a

situation and total extension referrable to with a projectible predicate (see the condition on p.37) can be seen as the nucleus of the presuppositions we have concerning *all* projectible predicates. It can be understood as a sort of general rule for use for predicates employed in practices oriented by the presumption of the possibility of learning from experience. As such it is (as Goodman saw) part of the background knowledge (the suppression of which gave rise to the riddle), and additionally as a *normative presupposition* about the efficiency (with respect to the aim of learning) of a given classification.

This nuclear presupposition has two components, which are of metasemantic and cognittively reflexive character (i.e. show that to be able to use projectible predicates, one has to suppose a minimum of distance to one's home langage), a quasi-ontological and a meta-epistemical part.

On the one hand there is the quasi-ontological assumption that the things falling under a concept have a common trait (are, relative to some background categorization, to be the same if numerically different) which «justifies» the classification independently of the concrete method of identification of individuals *as* falling under the concept. It is metasemantic in the following sense: assuming this, we simply count only domains (or models) as admissible that satisfy this assumption, make it, in other words *come out true*. It is, like in the case of genuine names or natural kind terms, an implicit restriction of admissible interpretations[41].

On the other hand there is the meta-epistemic assumption that the substantial, epistemically operative knowledge associated with the predicates that permits their identification *as* members of the extension is subject to continuous change (and does, therefore, never amount to a necessary and sufficient condition to determine *the* extension — because corruption is always possible, be it by us, by our errors or our world), whereas *this alone* does not bring about substantial changes in the domain.

The first part accounts for the preconditions of a referentially determinate use (which is thus construed as a supposition and so no direct negation of underdetermination but rather a strategy to cope with it) and the second part accounts for the openness and variability of the determination whether a predicate has been «correctly applied».

5.—. Some similarities

fine and very clarifying account of the structure of the argument see Hallett, M.: «Putnam and the Skolem Paradox», in: Clark, P./Hale, B. (eds.): *Readiing Putnam*, Oxford 1994, pp.66-97.

[41] It is never total, though: we can, with a change in «colour theory», most probably expect that our former predicates for colours in general will all be corrupted with the condition «Something is of colour x iff of colour x until the new theory was accepted or of colour y afterwards» or something like that. To suppose projectibility in a *continuous* sense we should then have to wait for a «unified colour theory». But in this section I am talking about the presuppositions we make from *within* an induction-based practice, and this is *essential* for the acceptability of realist assumptions, I think.

After having looked at the two problem-clusters of projectibility and kind terms, I want to stress some of the similarities that seem to be central to the use of general terms in either case.

The first and most striking similarity is the fact that gives rise to the problems: underdetermination of reference by interpretation. To take the problems discovered by Goodman and the natural kind theorists seriously is to accept from the outset that empirically interpreted predicates are not equivalent or coextensional to some one description of their extension where they do not occur. Nevertheless in both cases the presumption of referential transparence is central to the practices that would get in trouble were these problems operative. A theory of what we do when we suppose it, i.e. what background assumption falls if we should *discover* that some terms naively used as if referentially transparent are not really so (cases in medicine abund, but even a rise in differentiation in measure theory or an unnoticed extension of paradigms can prompt such a discovery): namely the assumption that the things falling under a kind concept are not really of a kind, put in terms of the practices: that the kind term is none and thus our generalizations might partly or generally be mistaken, such a theory is therefore an intrinsically pragmatic theory and no enterprise in metaphysical ontology.

This is the second similarity that seems remarkable to me: in both cases the assumption of referential transparence in absence of complete knowledge of the extension has the status of a *rationality assumption*. In that sense we could agree for both cases with what Føllesdal formulates for the first case thus: «Sameness of reference is *never* guaranteed.» (loc.cit., 110) The assumption of referential transparence or continuity is in that sense *not strictly epistemic or normative* and unfundamentable in the sense of not being logically or otherwise deducible. Accordingly Føllesdal goes on saying: «I look upon rigidity as an *ideal*, (...) that *prescribes* the way we use language *to speak about the world*. (...) All our talk about change, about causation, ethics and knowledge and belief (...) *presupposes that we can keep our singular terms referring to the same objects*. To the extent that we fail, *these notions* become incoherent.» (ibid., 111) Nothing to add except the stress on the fact that learning from experience is one case of «change, knowledge and belief» and that therefore, if Føllesdal's conjecture is right (and valid for general terms, as I hope to have argued), also this concept gets mysterious if we do not provide an adequate account of reference that explains the cases where we do *not* (to our knowledge) fail.

The analogy between Goodman's problem and the problems Witgenstein treats in the *Philosophical Investigations* (addition, pillar) that has been stressed by Kripke[42] and both to the problems that prompted the reflexions on the foundations of

[42] This is, as mentioned before, a hint to the *normative character* of a possible reconstruction why we do not always get confused by Goodmanian predicates. The elucidation asked for is, as far as I can see, how to make clear *why* corrupted predicates *must not* occur in certain practices, and not a general answer to the sceptic, that is, a discovery of something that is the case that *makes them not occur in fact*. An assumption raised by such a reflexion is of *intrinsic normative* character because for the reasons given it is neither plausible nor even desirable to exclude Goodmanian predicates *a priori*. The only answer to the problem raised by corrupted predicates asked

reference (more general: interpretation-) theory in form of accounts of 'direct' reference suggest that the operative assumptions, even though they have ontological and epistemological import, are interpretable as formal conditions of a certain manner of use of general terms (perhaps the «referential» use?). They do not constitute an *empirical* knowledge of certain facts or properties of the world: to presuppose projectibility or fixedness of reference through representative samples is not really to have learned something *about the world* for to be able to do *that* we have already to have projectible predicates. It means rather to have learned something about the relation between language and world, to have learned to differentiate by way of reflexion between language and the world described by it. This is the third similarity I see: that in both cases we get aware of some *reflexive capacity*, namely the capacity not always to confuse the result of given identification procedures (and «operational definitions») with *the* reference, the linguistic categories with confirmable structures in the world etc.

The fourth similarity one can extract from what has been said so far, if it is not completely erroneous, is a strikingly Kantian consequence (which, however, was already foreseen by Goodman[43]). It concerns the epistemological status of the assumptions having to be taken for granted if we assume projectibility and/or counterfactual substitutivity or fixedness of reference. Both assumptions are *a priori* relative to a practice of application of the respective predicates but nevertheless only

---

for is one from within the field where they actually cause trouble. If that is right, then the only thing we have to make clear is what exactly we do *if* we exclude them and what are the assumptions we have to make to be able to do so. This is the structure of the answer to the question how it is that we do not always stumble over corrupted predicates and consequently err in our inductive behaviour. These assumptions may have ontological import of a general sort (like the differentiation between sign, interpretation and object) and,, in virtue of that structure, exclude certain interpretative strategies as unapt to serve this aim (render certain interpretation theories wrong for an account of this behaviour and the contribution of language to the success of general behaviour), but one must not forget that this does not «prove» them to be «true». They are part of a rationality strategy *seen from inside*. From a participant perspective in the mentioned practices we certainly assume the existence and independence of our objects of investigation from the outcomes of the investigation, that is, we are and have to be «internal realists». However, this does not in the least mean that the ontology supposed in these practices has to be seen as any more priviledged than are these practices themselves in our conception of ourselves. This is to my mind the reason for the steady insistence on «explanatory relevance» of a common trait, for this is a case where the privilege of being worth to be pursued — explaining, that is — is almost too evident to be stressed. Especially it does not justify a claim to the effect that this is *the* world and the normativity integral to these assumptions has not to be misunderstood in the sense that, biewed from the outside (possibilitated by e.g. an alternative account of the domain) we have to hold on stubbornly to some set of categories.

[43] Kripke, S.A.: *Wittgenstein on Rules and Private Language*, Cambridge MA 1982, esp. p.20.

to be motivated, specialized for each single predicate and to be tested *a posteriori*. So the two assumptions, conceived of as two aspects of a capacity to distinguish world and language (for each language, but not, of course for all languages), are *synthetic a priori*, where the «*a priori*» is, obviously sort of *contextual*. Kripke calls the respective assumption for modally stable terms, as is well known, «*a posteriori* necessary». This can be given the following, «deflationist» reading: by this selection he stresses the sort of non-analytic but strict validity that we impute on the rules for applying these predicates: these rules serve as a standard for admissible interpretations (and thus are, from a purely semantic point of view, rendering a concept of «necessity») as long as we consider the generalizations articulated in them of explanatory force or whatever worth, and this evaluating, reflexive activity is not to be accomplished by logical truth or analyticity. Kind words exist in this view thanks to the experiences we make with things in the world and only assuming them to be such we can procede to an investigation of the objects in the domain that *must not* substantially change *because of the results of the investigation (the changing descriptions) itself*. The experiences which prompt the generation and are involved in the introduction of kind terms are made with arbitrary objects or «contingently» given «samples», where their *being samples of a kind* is, again, an *a priori* assumption concerning the homogeneity. Thus the set of introduction-paradigms is «*a priori* contingent». In both formulations we can thus see substitutes for the *synthetic aprioricity* of the presupposistions that are inevitable for the generation of kind words. Of course, Kripke would probably charge this treatment of undue *epistemologization of metaphysical categories*; but if we do not pronounce any opinion as to their place in theories (which is the same as relating epistemology and (meta-)semantics), these categories become quite pointless. In sum, I think that one can say that all of Kripke's metaphysical conclusions are only insofar essential to an explanation of the behaviour of expressions that are flexible enough to cope with changes in our knowledge without being unduly flexible in their reference as they can be reconstructed as indications of *normative* conditions of the use and interpretation of predicates within inductive practices. One of the best expressions for this way to connect the heavily charged notion of «necessity» with the *reflexive attitude* needed for this manner of use can be found in Donnellan's early article «Necessity and Criteria»[44]: «Whether [a determinate statement, e.g. one that relates a property considered as important for the generation of a kind, an «underlying trait» and a given predicate, A.M.] *is*, as we *intend* it, a necessary truth or contingent, is *indeterminate. It is indeterminate because the* ***decision*** *as to which it is would depend on our being able to say now what we should say about certain hypothetical cases*. (...) Necessity (...) might be thought of as an *ideal rigidity in our* ***judgments*** *about what to say concerning hypothetical cases*.» (S.658)

6.— Speculations on the relations between projectible and natural kind terms

[44] Cf. FFF, p. 96: «Somewhat like Kant, we are saying that inductive validity depends not only on what is presented but also upon how it is organized; but the organization we point to is effected by the use of language and is not attributed to anything inevitable or immutable in the nature of human cognition.»

The question of the exact relation of projectible to natural kind terms remains open. An answer to it will depend on the specific account in which the behaviour of both types of predicates and related ones (e.g. dispositional predicates) will be described. I will only adventure one hypothesis in this respect. In general the conditions for being treated as projectible should coincide more or less with some versions of what Føllesdal said about the motivations for introducing names. It seems probable that what is intended with the classification of some general terms as natural kind terms is somewhat more specific than what is intended by a qualification of predicates as projectible. Inbetween I would expect the dispositional predicates. Thus the relations would be: not all projectible predicates are apt to constitute natural kinds, but for a given predicate to be a natural kind term it is inevitable that it be projectible. Projectibility would then be a necessary but not sufficient condition for being a natural kind term. The methodological priority which is often given to the latter is probably the consequence of the fact that they underlie our most common classificatory practices. On the other hand one can, as I have tried to make clear, learn something about what projectibility consists in through an analysis of the rationality-presuppositions involved in the use of these so common terms.

It does not seem all too far fetched to suspect that terms that are treated as projectible are natural kind terms iff they occur as fundamental concepts in theories of natural science (as opposed to social science and others).

The class of dispositional predicates seems also to be more general than the class intended by the term «natural kind terms», but it is an open question if all natural kind terms are to be analyzed as disposition terms. However it seems conceivable to me that this is so in view of the fact that what is done in natural kind term theory is to establish a relation between underlying, unknown and known, superficial properties of objects, which is exactly what one does when imputing a disposition to some object. Dispositions are, however, less firmly linked to theories about facts of the objective world and preferably excluded by them as *explicit* dispositional predicates. In that case, natural kind terms would be the accepted correlates of disposition predicates for natural science. This might sound a bit strange at first sight, but the decisive point that has always been made to differentiate natural kind terms from n-criterion words is clearly that in the case of natural kind terms there is, in addition to some manifest community, an (causal, microstructural or whatever) *explanation* how this community is brought about, although this is something we (can) have only *indirect* knowledge of as long as this explanatory trait is only accessible by the analysis of some *testable* manifest traits and the reaction of the things that have them. On the other hand dispositional predicates also have some traits that resemble names, as has been claimed for natural kind terms: they are descriptionally inexhaustible and help us to generate sets of things, all of whose members we refer to by calling them e.g. «intelligent», «soluble», «being one meter large». Thus the structure of application is *implicitly* dispositional, I am inclined to suspect (both types are, to remind of an almost forgotten attempt to treat this kind of questions, introduced by some sort of «bilateral reduction sentences» and afterwards used *as if* they were «normal» predicates although it is known that they are not defined and they are kept as reference-constant through changing operational conditions to determine membership (this is the *advantage* of not being defined but being, nevertheless, accepted as referring to some explanatory relevant grouping of things)).

Dispositional predicates do also have to be projectible, but in contrast to the completely general supposition of an existing homogeneity there is, in the case of dispositional predicates, an explicitly stated criterion for the decision whether it is justified or not. This might be expected to be found in all projectible predicates, thus perhaps both classes coincide under the condition that the projectible predicates are to be interpreted empirically. But these remarks are, I want to stress, by now merely speculative.

7.— Summary and Conclusion

The specificity of natural kind terms seems to be that they are our means to constitute domains of investigation. We could call them synthetic categories. In what sense they are dependent on purely formal, essentially synthetic but nevertheless contextually a priori presuppositions can be seen when they are viewed as a special case of projectible predicates.

I think that the behaviour of natural kind terms and our behaviour using them show in an exemplary way a specific formation of ontological and epistemological background convictions concerning the relation of language, our use of it, and reality. That the description of the rules for their use consists simultaneously in a description of the rules for predicates apt to be used in inductive procedures implies that a part of these convictions concerns deeply our relation to past experiences and expectations about future experiences with reality. The presuppositions for the use for predicates usable in induction are obviously at the same time the ones needed for the possibility of structured learning processes. So the reflexion on the conditions for the use of natural kind terms, which have, as we saw ontological and epistemological import, can apport (some of) the philosophical assumptions taken for granted in the talk of «learning from experience» or, to put it differently, what the assumptions are one is committed to when adopting a cognitivist attitude towards our experience with the world. As soon as an agent supposes to learn from experience, he has to accept some version of the presuppositions (or more) indicated above; they are part of the general background knowledge that makes possible that we deal in an ordered way with past experience and access to some such way to evaluate new ones.

The question of how it is possible or better: what it is to apply a kind word in a determinate way is answered by the theory of reference with the seeming triviality that this is the case iff we always refer with it to the same: all individuals of a kind. The mentioned question reminds undoubtedly of Wittgenstein’s incessant questions on following a rule. Now, taking this reminder in account, one could say that the exciting «discovery» in the course of the work on a theory of the interpretation of natural kind terms was exactly to show that the *specific theory* of rule-following for that case has inevitably to be a theory of reference and is not possibly substitutable by any account based on meaning that consists in the attribution of some set of deterministically conceived, substantial rules that function according to the example of analytical or logical truths. This has been resumed by Putnam in the mouthshell that in the case of the interpretation of this sort of terms «reference does all the work»[45]. The strictness of the validity or «necessity» of the rules for the interpretation of natural kind terms is not exactly analogous to logical truth; it is not primarily due to our relation to the intersubjective undisputability of logical truth but rather to our relation to experiences in the objective world and our conviction

[45] In: *Journal of Philosophy* 59 (1962), S.647-58.

articulated in it that the world is independent from the beliefs we maintain *de facto* (although it is not, of course, independent from experience and language in general: rather every use of language articulating our experience presupposes necessarily some object of experience). This discovery of a «non-analytic necessity», as Deutsch[46] puts it, is, from my point of view, the most important result of the so called theory of «direct» reference and is, thanks to its general character reconstructible and obtainable without most of the fundamentalistic metaphysical convictions associated with a good deal of the work done in this area[47].

In sum, the (semantical, epistemological, pragmatic,, ontological) differentiations between sign and signified, reality and construction, reference and transmission of what is meant and, above all, our capacity to draw them, seem to be unseparably linked to the cognitive inventory that we invoke when we talk of «learning from experience», «the independence of confirmation instances» and the like.

Thus any theory that blurs these differentiations is incompatible with a claim to the effect of the possibility of learning, improving theories etc. A deterministic theory of reference that tries to reduce the reference of the terms to a mechanism between the factual substantial knowledge associated with the term (its meaning or one determinate description of the extension) and objects that satisfy this knowledge is incapable of describing adequatly the behaviour of the participants in practices who assume them to serve the aim of learning. To attribute them a capacity to learn and criticise *de facto* existing beliefs and a cognitive attitude towards hypotheses is incompatible with describing their interpretative behaviour with a deterministic theory of language.

Axel Mueller

Frankfurt University

amueller@stud.uni-frankfurt.de

[46] Cf. Putnam, Hilary: *Representation and Reality* (Cambridge MA 1988), S.46.

[47] «semantics for Natural Kind Terms»