



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2005

3 WD meets GTVH: Breaking the ground for interdisciplinary humor research

Hempelmann, Christian F ; Ruch, Willibald

DOI: <https://doi.org/10.1515/humr.2005.18.4.353>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-77746>

Journal Article

Published Version

Originally published at:

Hempelmann, Christian F; Ruch, Willibald (2005). 3 WD meets GTVH: Breaking the ground for interdisciplinary humor research. *HUMOR: International Journal of Humor Research*, 18(4):353-387.

DOI: <https://doi.org/10.1515/humr.2005.18.4.353>

3 WD meets GTVH: Breaking the ground for interdisciplinary humor research

CHRISTIAN F. HEMPELMANN and WILLIBALD RUCH

Abstract

The present paper describes an interdisciplinary effort, in which results based on the same material, but analyzed with tools from two different disciplines are brought together for mutual evaluation. The set of 70 jokes and cartoons from the 3 WD (Ruch 1995), which has been extensively studied psychologically for its affective properties, is analyzed linguistically for its internal morphology based on the General Theory of Verbal Humor—GTVH (Attardo and Raskin 1991). The correlations between the stimulus properties and their effects are discussed, as well as the relevance of these results for the respective theories and the disciplines that use them. Additional emphasis is placed on highlighting the problems and considerable benefits of such interdisciplinary research as the most apt approach to complex phenomena like humor. The results show that there is indeed significant overlap between stimulus properties as they can be distinguished linguistically and affective responses as they can be identified psychologically. Of the six GTVH categories, it is primarily script opposition, narrative structure, target, and logical mechanism that contribute to the separation of the three humor types with respect to effects on recipients. The results also suggest that initial and residual incongruity, as operationalized with the GTVH, are central cognitive aspects of humor with an impact on affective factors and, consequently, their distinction. While this may appear to be commonsensical results, their scientific reproduction is a major step forward, in this case for humor research.

Keywords: 3 WD; GTVH; humor appreciation; interdisciplinary research; linguistics; psychology.

1. Introduction

This paper attempts to address a desideratum of humor research that is as old as its revival in the 1970s: namely, interdisciplinary cooperation in a field that requires it more than many others (cf. Apte 1988, Raskin 1995). The two disciplines whose theories and methodologies are brought together here are among the main motors in current humor research. Their cooperation has long been envisaged and previously realized in Ruch et al. (1993). Ruch and Hehl (1998: 138f) posit the obvious desideratum to which the present article is the response: “Once a pool of jokes varying on all dimensions and preclassified on [the GTVH] parameters is available, confirmatory factor analysis could be applied to derive empirical weights for the relevance of the different modes. A failure to verify the importance of one mode (for the ordinary recipient of jokes) means that this knowledge resource does not affect *differential* appreciation of humor; however it does not speak against the theoretical significance of that knowledge resource in the morphology of jokes.”

We assume a general familiarity with the two approaches: linguistic humor research in the vein of the General Theory of Verbal Humor (GTVH) and the psychological humor research based on the 3 WD questionnaire. While general information about these can be found in several sources (Attardo and Raskin 1991; Raskin 1985; Ruch 1992; Ruch and Hehl 1998), it is crucial to clarify the different positions from which the two strands of research start out and their related, yet different, goals.

2. Theories, methods, and goals

The Semantic Script Theory of Humor (SSTH; Raskin 1985) posits that “a text can be characterized as a single-joke-carrying text if both of the [following] conditions are satisfied: (i) The text is compatible, fully or in part, with two different scripts (ii) The two scripts with which the text is compatible are opposite” (1985: 99). The SSTH carefully develops a formal-theoretical basis for the core concept of *script*, and it provides a methodology for script identification and analysis. Simply put, a script is “a cognitive structure internalized by the native speaker and it represents the native speaker’s knowledge of a small part of the world” (Raskin 1985: 81). A recent expansion of the formalized notion in semantic theory can be found in Nirenburg and Raskin (2004, cf. also Raskin et al. 2003).

Attardo and Raskin's (1991) general linguistic revision of the SSTH encompasses six knowledge resources (KRs) ordered hierarchically: script opposition (SO), logical mechanism (LM), situation (SI), target (TA), narrative strategy (NS), and language (LA). The hierarchy of KRs was empirically tested on joke similarity (Ruch et al. 1993). Recent work has extended the theoretical basis of the GTVH, in particular the role of the LM (Attardo et al. 2002), the relation of script oppositeness (SO_p) and script overlap (SO_v; cf. Hempelmann 2004), and its relation to cognitive processes (Attardo 1997, Brône and Feyaerts 2004). Thus, the SO of the GTVH aims to identify the necessary and sufficient elements of a text for it to be processed as humorous, while the SO and LM elements in particular model the semantic-cognitive processes involved in this specific type of making sense. The GTVH at large, encompassing these and the four further KRs, is a joke representation model intended to distinguish variants and invariants of jokes as the most representative subset of verbal humor. It is neutral to affective response factors beyond the assessment of the humorous status of a text and can serve as a tool to study textual humor wherever the research question needs to be addressed by the analysis of the stimulus.

In contrast to the GTVH, the taxonomy of humor underlying the 3 WD humor test (Ruch 1980, 1992) is a two-mode model of humor appreciation as perceived by the layperson and systematized by a personality researcher. With respect to the stimulus mode, the 3 WD distinguishes three categories labeled—not analyzed from a linguistic point of view—as incongruity–resolution, nonsense, and sexual humor. The 3 WD also separates positive (funniness) and negative components (aversiveness) of humor appreciation with respect to the response mode. These categories were derived with the help of factor analysis, a mathematical-statistical tool used to empirically develop taxonomies.

The research agenda in factor-analytic studies is to first define a universe of items (e.g., jokes and cartoons that can be presented in experiments). Then a rule is generated that allows drawing a representative but manageable set of, for example, 100 jokes and cartoons from that universe to be used in further studies. While they may be theoretically analyzed or pre-classified by a number of criteria, the essential classification is done empirically, utilizing a large number of participants (e.g., 300 adults who are representative of a population). In such a study, each individual will be presented each joke/cartoon and his or her responses, e.g., items rated for degree of funniness, are recorded. The application of factor

analysis then yields an initial taxonomy that requires interpretation. Factor analysis yields the number of humor categories and which jokes belong to what category. Determining the nature of these categories is a subjective step in which the researcher attempts to find the common aspect of jokes in one factor, typically using the *a priori* analysis of the jokes. Once this initial taxonomy is derived, replication studies are conducted substituting both the sample of subjects and sample of jokes (maintaining the labels of categories). Such studies might yield slightly different or additional categories and the model and/or interpretation of the factors might shift. Once a satisfactory stage is reached, the best markers of the factors are selected and used to construct a test for future standardized assessment of the constructs derived, here: the factors of humor appreciation.

In the case of the 3 WD, the three emerging factors had to be elaborated only slightly (McGhee et al. 1990; Ruch 1980, 1992) after their initial interpretation (Ruch 1980), which remained consistent in replication and validation studies. Note that the interpretations of the factors are not mere description of the material, but combine the description of the jokes' features with the description of the peculiarities of their processing. McGhee et al. (1990) describe the salient features of incongruity-resolution (INC-RES) humor, one of the three factors, as being "characterized by punch lines in which the surprising incongruity can be completely resolved." With respect to nonsense (NON) humor, the second factor, they write:

[N]onsense humor [...] also has a surprising or incongruous punch line, exactly as in incongruity-resolution humor. However, '... the punch line may 1) provide no resolution at all, 2) provide a partial resolution (leaving an essential part of the incongruity unresolved), or 3) actually create new absurdities or incongruities.' (McGhee et al. 1990: 124)

In nonsense humor the resolution information gives the appearance of making sense out of incongruities without actually doing so. (see also Rothbart and Pien 1977)

We would like to note here that with respect to full, partial, and no resolution, we take a different and more careful position: Not least because of the present results, we assume resolution in humor to always be partial, as the logic that enables it is always playful or faulty. Thus, incongruity-resolution humor should be considered one extreme: namely, one closest

to, yet distinct from, full resolution. Nonsense humor, on the other hand, takes up the opposite extreme: it is closest to a complete lack of resolution, but it still appears to have one. This latter position corresponds closely to that of Rothbart and Pien (1977). Understood in this way, the interpretation of the 3 WD allows for the two dimensions, INC-RES and NON, to be construed as related extremes of a continuum of cognitive processes. We will return to this point in connection with the elaboration of the semantic KRs of the GTVH that model these cognitive processes.

Both INC-RES and the NON structure can be the basis for harmless as well as tendentious content, as in the case of sexual humor. The sexual humor factor (SEX) was initially the easiest to identify due to its salient content. Furthermore, it was the only one of the three factors that was expected to appear, since a factor of sexual humor has been found in almost all previous factor analytic studies. Subsequently, however, it was discovered that SEX jokes and cartoons typically have two loadings: one on the SEX humor factor, as well as a second one on one of the two structural factors, INC-RES and NON. The size of this second SEX loading seems to depend on the degree of the theme's salience. In very explicit items (mostly cartoons) the loading on the structural factor is very low, whereas in less salient items the loadings on the content and structural factor can be of roughly equal size. According to their loading patterns, the items of the general SEX humor category can be subdivided into three classes: "pure" SEX humor (in which the content largely overpowers the structure), INC-RES-based SEX humor, and NON-based sexual humor.

The 3 WD test of humor appreciation was applied for several of the purposes described in Ruch (1992: 27): e.g., as a test of humor appreciation; a means of testing the universality of humor categories (Ruch and Hehl 1998); a tool testing Freudian theories of sexual drives and humor appreciation (Ruch and Hehl 1987); a standardized means to induce amusement (Ruch 1995). In addition, because strong relationships were found with selected personality variables, it was discussed whether it can be used as an 'objective' test of personality (Ruch and Hehl 1985).

The objective of such a classification is related to, but different from, classifications in other disciplines. For example, differences within a joke category are *not* of interest; on the contrary, the shared element is essential. While the system does allocate different classifications to different jokes (the factor loadings are used as coordinates to locate the jokes in the factor space), the differences among jokes are quantitative (i.e., dependent on how much they are influenced by the different factors), not

qualitative. Furthermore, the goal is to get “pure” representatives that retain the markers for those factors, while jokes that have relations to more than one factor—i.e., jokes that lie in between factors in the factor space—are eliminated. This is justified, because the factors are intended to contain the essential information, while the jokes loaded for several factors can be emulated by mixtures of factors. Thus, the ultimate goal in such a research project is to have a set of categories of jokes/groups that function in the same way, each of them represented by a set of markers, as purely as possible. In fact, the jokes of one category are summed up to form a total score of liking of that humor type, thereby strengthening the variance common to those jokes and averaging out irrelevant ingredients.

In contrast to this, a linguistic classification of jokes might be interested in assigning *all* jokes a separate classification, in which there is only *one* joke per cell of a taxonomy. In this sense the 3 WD classification is reductionistic and focuses on what is believed to be the “core ingredient.” Therefore, the application of a linguistic theory like the GTVH allows for a rich description of any textual corpus studied for a variety of possible research questions. Of course, it may be the case that not all of the features identified by such rich description are actually relevant to the responses of lay people.

3. Interdisciplinary work

In the previous discussion, a fundamental difference between a linguistic and a psychological approaches has emerged. Linguists analyze elements of the text and its cognitive processing from a theoretical *expert* perspective intended to emulate an idealized language user. The psychological approach utilized in the 3 WD studies, on the other hand, examines the various responses to the jokes and cartoons by a naive *recipient*. The psychological taxonomic studies identify *types* of humor based on similar mechanisms leading to similar degrees of funniness. Focusing on humor competence rather than performance, linguistic studies primarily ask, “What is a joke?” and not “How funny is a joke?” It might well be that some features distinguish jokes from non-jokes, but other parameters that affect the degree of perceived funniness, possibly more strongly, e.g. humor-related traits or contextual and aesthetic factors of the joke presentation are not focused on (cf. Samson and Hempelmann, n.d.).

In other words, in contrast to the aims of a psychologist, a linguist uses a theory of joke texts to determine whether a certain stimulus is a joke and what elements are necessary and sufficient for the joke to contain humor. It is a different, if related, task to predict the impact of the text/cartoon on a recipient beyond his or her identification of it as a joke. This aspect crucially holds for a prediction of the perceived funniness of the text, which is influenced by many individual psychological factors of the recipient. The most that can be achieved is to use the theory of joke texts to categorize the texts according to specific parts theoretically predicted to be essential, and to look for a correlation between those categories and patterns of recipient appreciation of these texts. If the categorization is done independently of the recipient testing and such correlations can be found, they may help in the identification of stimulus factors that account for the variations in the recipient appreciation. This is our rationale here.

In general, the relation between joke ingredients as identified by linguistic analysis and perceived funniness can be analyzed in different ways. First, one may attempt to link variation in those ingredients to perceived funniness in a general statement. The variation may be achieved experimentally by manipulating existing jokes so as to either contain or not contain a certain ingredient (cf. Ruch et al. 1993). The variation may also be achieved by changing the degree of the ingredient, or quasi-experimentally, as in the present case, by selecting different jokes that already contain or do not contain the ingredient, or do so to varying different degrees. On this basis, the effect of the variation of this ingredient on perceived funniness can be examined. Or to put it in terms of the GTVH: Can we predict which KRs will have differential effects; e.g., are those higher or lower in the hierarchy SO, LM, SI, TA, NS, LA, more influential, or the ones related to content (SO, SI, TA) vs. those related to form (LM, NS, LA)? This question might, for example, yield the result that the degree of incongruity affects funniness in a way described by an inverted-U-shaped curve: Low and high degrees of incongruity lead to lower funniness, while medium incongruity is optimal for funniness (cf. Giora 2002). Such a finding is assumed to be *generally* valid (for all persons) and at the same time one makes comparisons across stimuli.

Second, one may focus on differences in humor-related traits. Here, the ingredients do affect funniness too, but this relationship is moderated by personality. There may still be an inverted-U relationship between degree of incongruity and perceived funniness, but, for example, the location of an 'optimal' incongruity may be different for people who in general are

more tolerant or phobic toward incongruity. Here the comparison is across individuals, and the 3 WD studies are built on the latter assumption: People are different *a priori*, and therefore different types of humor have developed to match the different aesthetic preferences, which might not be specific for humor, but represent more general preferences, independent from the domain of aesthetic experience. Humor can therefore also be taxonomized on this basis and the relation to other aesthetic preferences, or—more generally—personality dimensions, can be explored.

In sum, the aim of the present study is to reassess the properties of the three humor categories derived by factor analysis—this time from the point of view of linguistics. In order to achieve such results, the 3 WD stimuli are analyzed linguistically based on an expansion of the GTVH described in the next section.

4. Expansion of the GTVH

The GTVH had to be adapted in several ways in order to be applicable for the present task. This task is unusual—and very problematic—because it does not directly address a question from the target field of application, here humor, but is theory-driven. Such a perspective is unavoidable for the evaluation of theories, such as the linguistic humor theories from psychology and humor under discussion here. This entails practical problems in that there is no legitimized level of abstraction for the application of the GTVH, because where a question is lacking, the format of an answer remains in principle indefinable. To counter this problem, we chose to expand the descriptiveness of the most crucial higher-level KRs, SO and LM, sharing the general assumption that “these two will turn out to be much more useful for generalizations in humor research than some accidental assortment of lower level arguments” (Attardo and Raskin 1991: 329). While this will be addressed in detail in the next subsection in the hope that a discussion on these important issues can be initiated, another expansion will not be addressed explicitly in the present paper: The 3 WD relies heavily on cartoon humor, which uses pictorial stimuli or a combination of pictorial and textual elements. The implicit assumption in previous research (e.g., Paolillo 1998) has been that the cognitive-semantic processes involved in cartoon humor are generally compatible to those claimed for purely textual humor by the GTVH. We share this assumption of compatibility of the processes, at

least for the factors relevant here, but will leave a detailed discussion of differences and problems, in particular in relation to the different aesthetic dimension of cartoons, for later (cf. Samson and Hempelmann, n.d.).

Thus, it should be noted that the present section focuses on the expansion of the GTVH. As such, it goes beyond the description of methodology one would customarily expect in psychological papers. It is rather an important theoretical development that forms the basis of the expert rating of the stimuli.

4.1. *Script opposition and degrees of incongruity*

A central theoretical problem that had to be approached for the present task was the quantification of incongruity, as well as resolution, as conceptualized on the basis of the GTVH in the KR_s SO and LM, respectively. This entails the development of a methodology to determine this degree in a reproducible fashion—as Attardo et al. have pointed out: “It may be interesting to investigate whether an increased level of resolution is reflected in higher appreciation by speakers, and, more in general, if the degree of resolution is somehow connected to other factors in the joke (for example, to the degree of incongruity)” (2002: 26). This will be partially attempted here, as it is necessary to arrive at data that can be correlated in the fashion we require.

One obvious possibility to conceptualize degrees of incongruity is in terms of a notion of “distance” of scripts or “Fallhöhe” (Gernhardt 1988). But this spatial metaphor is useless until it is connected to a theory that places scripts at “distances” and provides a methodological tool for measuring different script “distances.” Our concept of scripts is based on ontological semantics (Nirenburg and Raskin 2004). Thus, the available tools of *Ontosearch*, developed within this paradigm (Onyshkevych 1997) and implemented for the *Mikrokosmos* system (Mahesh et al. 1997a, Mahesh et al. 1997b), could be emulated for a small number of local analyses. But scripts in jokes are instantiations of scripts, with foregrounding, specific slot-fillers, and salient gaps (cf. Attardo et al. 2002; Brône and Feyaerts 2004) whose opposition holds “only within a particular discourse and solely for the purpose of the discourse” (Raskin 1985: 99; cf. Attardo 1997 on ‘local antonymy’). Thus, the specific oppositeness relation between two of them in a joke is not simply the number and

types of concepts in an ontology that one has to traverse to get from one to the other, as *Ontosearch* does. Rather, it is a specific relation of those two instantiations of scripts in this very text (for the very audience), with various degrees of filled and salient slots in the scripts. For a measurement of degree, we need to identify the type of oppositeness, and, more importantly, we need to classify these types into a theoretically motivated hierarchy representing the degrees of incongruity of these types. Again, we'd like to caution that this task tries to achieve something that goes beyond what the SO and LM of the GTVH were originally designed for. It will be based on the GTVH in crucial aspects, but its faults in its current tentative state remain our responsibility.

One candidate for a hierarchy that could form the basis for such a measure of incongruity is the (open) set of abstract SOs: “[M]any jokes evoke one of the relatively few binary categories that are essential to human life” (Raskin 1985: 113), e.g., *sex/non-sex*, *life/death*. The problem here is that we lack a criterion to rank these categories and that developing one is a task beyond the scope of the present research, as it would involve the analysis of the overall importance of these issues for human lives. In order to utilize these categories, we will rate the stimuli in accordance with them under the category ‘SO concrete’ without assuming a ranking in relation to degrees of incongruity.

Another non-implementable approach would base the degree of incongruity solely on the degree of script overlap of the joke text, namely *full*, *partial*, or *truly partial* (Raskin 1985: 105f). This overlap is also local for each joke, and we definitely assume it to exert an influence on the joke's quality and perceived funniness. But the overlap ignores the *type* of incongruity that should clearly dominate in a concept of incongruity degree. Thus, a combination of factors of type and instantiation of scripts may be the better avenue for a future approach.

An existing concept that does unite the local constellation of scripts in a joke text with the general type of their oppositeness is the set of *un/real* relations into which two scripts are brought in a joke (Raskin 1985: 127):¹ *im/possible*, *ab/normal* (un/expected), or *non/actual*. For our present purposes, there are at least two problems: (1) there is, again, no existing hierarchy among these three abstract SOp types, but we think a ranking is implicit (see below); (2) in many jokes, there is “shadow oppositeness” (Raskin 1985: 108), e.g., many instances of non-actual/actual oppositeness are shadowed by an actual/non-actual one. We will ignore the second point for the same reason that we ignore secondary SOs in

stimuli: namely, simplicity. Rather, we will focus on the development and justification of the more important first one.

The main motivations for our hypothetical hierarchy of SOPs are reflected in the way we would rate a joke for type of oppositeness: Only if both scripts are possible, is it meaningful to ask whether the second one is abnormal or non-actual, because an impossible script is abnormal in its impossibility and can only be presented as non-actual, e.g., *elephants sitting on marshmallows* or *cookies crying*, neither of which is possible. Similarly, Rothbart and Pien (1977) assume the highest degree of incongruity to hold for impossible incongruity. Actuality and normality of opposite scripts, on the other hand, are not strictly dependent. Sailors actually do not necessarily use *dirty words* whenever they use *nautical terms*, as suggested by a pun on naughty vs. nautical (cf. Raskin 1985: 29). But it would not be abnormal or impossible. Consider the following pun (Raskin 1985: 26): *Should a person stir his coffee with his right hand or his left hand? Neither. He should use a spoon.* It is abnormal to stir one's hot beverage with one's hand. This abnormality implies that the act of stirring actually takes place, as the hypothetical voice of the joke's question indicates in the use of *should*. But stirring a hot beverage with one's hand is not impossible. In sum, we assume more incongruity to be present in the contrast between abnormal and normal scripts than between a script that is actually there and another one that is (revealed to be) not there (cf. Raskin 1985: 111), while both the of them may very well be possible and normal.

This hierarchy clearly neglects the influence of the specific un/real oppositeness in terms of *life/death*, *good/bad*, *high/low*, etc., but for the reasons outlined above, we assume that it forms a useful basis for the present task. So for the purpose of the present study the analysis of abstract SO in terms of non-/actual, un-/normal, and im-/possible is coded as follows (allowing one additional degree of freedom within the types to accommodate more extreme cases and rater variance in the numeric rating represented by boldfaced numbers):

- **4** higher/**3** high [im-/possible]
- **3** high/**2** low [ab-/normal]
- **2** low/**1** lower [non-/actual]

The following other factors were implemented or adapted from the GTVH for this study: As mentioned, *SO antonymy* describes which of the oppositions is the main opposition of the joke/cartoon. The rating

was open to additions, but the following list was supplied as a general guideline: *good/bad, life/death, non-/sex, non-/money, high/low stature, non-/feces*, etc. Thus, as its name implies, *SO antonymy* is a factor that describes the content of the *SO abstract* oppositeness relation. In addition, *SI/S2* provides a brief description of script one and script two of the main opposition, i.e., the main event/state that characterizes them. This was intended to provide an illustration of their relation very concretely and to check if raters agreed on the main scripts, the identification of which is the best starting point for a GTVH analysis.

The degree of resolution [*RES deg*] factor, derived from the type of LM, is aimed to capture the complexity of the false logic of the joke that connects the two scripts and thus, presumably, the effort necessary for its resolution. This cognitive effort is related to the initial degree of incongruity and possible additional incongruity as introduced by the resolution (cf. Rothbart and Pien 1977). There is a long tradition in research aiming to formalize the resolution of incongruity in humor, in the context of which Davies (2004) may be justified to note that mere taxonomies are futile products of “crass empiricism at best.” But we are convinced that analyzing the linguistic trigger, and the cognitive-semantic process of pseudo-logical resolution enabled by it, is a worthwhile undertaking leading to new insights and testable predictions. The test question here was, “How hard is it to comprehend the relation/overlap of the two scripts through the logical mechanism?” The possible answers range, again, in four steps from very simple to very complex with two intermediate values, resulting in a four-point scale. An obviously related parameter is the identification of the type of logical mechanism [*LM type*] for the main SO of the stimulus. A short description of the false logic that brings the two scripts into the overlapping relation and allows for the partial resolution had to be supplied. A list of possible answers was provided, including *pun (cratylistic analogy), false analogy, garden path, role reversal, shift of perspective*.

Also rated was the residual incongruity [*rINC*] of the joke that remains after its partial resolution through the LM or introduction of further incongruity. The test questions “how much incongruity remains un(re)solved; how much new incongruity has the LM introduced; how puzzled are you still at the end of the joke” could again be answered in one of four degrees, from lower to higher amount of residual incongruity.

Three further parameters represented the remaining knowledge resources of the GTVH with the exception of LA: *SI* (situation) describes

the main state/event described in the joke, often identical with script 1. *TA* (target) is the butt of the joke, if any. *NS/DS* (narrative structure/drawing style) captures the textual structure of the joke, e.g., narrative or question-answer, and, in the case of cartoons, the number of panels and the distribution of textual elements, if any, e.g., “3-panel cartoon with caption/heading.” The problematic nature of rating these KRs without the guidance of a research question has been discussed above.

For cartoons, the parameter *S1/S2/LM trigger* [script 1/script 2/script-switch trigger for the logical mechanism] specifies the location of these elements of the joke and was graded as either “pic(ture)” (with a number in case the cartoon had several panels) or “text.” It is not used here, but was included as a pilot measure for a separate study that will focus on the perception of cartoons as humorous stimuli, as this location may imply the processing order of text vs. picture(s). Another parameter used only for cartoons and not used here is their aesthetic quality [*cartoon aesthetics*].

On the basis of these ratings, the aim of the present study is to relate coded original (e.g., LM, TA, SO) as well as derived (e.g., degree of incongruity, degree of residual incongruity) GTVH parameters to empirically derived parameters (i.e., factor loadings, mean judgment of funniness) for the 70 jokes of the 3 WD humor test. This will allow for a more comprehensive description of the item pool of the 3 WD and supplement the interpretation given so far. Furthermore, the parameter will be used to examine how the three factors are different and thus enrich our understanding of INC-RES, NON, and SEX humor.

5. Method

5.1. Subjects

For the original factor analysis condition, the sample comprised 112 non-psychology students of the University of Düsseldorf who were paid for their participation. A principal component analysis of the 70 jokes and cartoons was performed. Three factors were extracted and rotated using the Varimax criterion. The rotated solution was used as a reference for the coded ingredients. While the factors were easily identifiable as INC-RES, NON, and SEX, the loadings might be slightly unstable due to the low number of subjects in this sample.

In the expert coding condition, there were 3 participants: the linguist co-author, an associated linguist, and an associated psychologist, all with postgraduate training in their fields in general, as well as humor research in particular. The linguists have worked with the theory that formed the basis for the analysis of the 70 stimuli for several years; the psychologist was trained to apply the criteria outlined above on the basis of her previous substantial exposure to the linguistic theory underlying it.

5.2. *Material*

The material for both conditions of the study was the 70 jokes and cartoons of the German version of the 3 WD A and B forms in their original booklet format. All 70 jokes and cartoons were analyzed. However, as the factor loadings of the warm-up items are unreliable, they were discarded from that analysis. Only 60 were considered for further processing in the factor analysis condition. No restrictions were applied for the expert rating condition.

5.3. *Procedure*

In the expert coding condition, the linguist co-author analyzed the 70 jokes with the criteria described in the design section. These are based on the adapted GTVH as outlined in the theoretical section. He was not aware of the clustering of the stimuli as it results from the existing 3 WD studies. The other linguist and a trained psychologist each analyzed a randomly chosen subset of 10 of the stimuli in order to produce data on the interrater reliability of the analyses. In both cases, their rating was preceded by a training session. The results and problems of the interrater reliability are discussed in detail below.

Where the co-raters differed insignificantly (e.g., in SO concrete, rater: money/non-money, co-rater: rich/poor), we adjusted toward the more general category, i.e., money/non-money, while indicating in parentheses the original overspecific choice. Where the co-rater used a category that can be subsumed under an available suggestion (e.g., “conversation” or “cartoon with 1 panel” for NS), we adjusted toward the closest possible suggestion—here “narrative,” resp. “1-panel cartoon.” Where the

co-rater made an obvious mistake, e.g., miscounting the number of panels, we quietly corrected.

5.4. *Interrater agreement*

For the expert coding of the KRs the percentage of agreement between the expert and co-raters was computed for each parameter across the 10 items. For the first co-rater, the average agreement was only 55% with the coefficients ranging from 20% (LM type) to 100% (narrative strategy). The numerical codings yielded similar results. Product moment correlations were utilized and were highest for judgment of residual incongruity where there was an interrater agreement of $r = .66$. The two other coefficients were lower but this might be explained by their reduced variance. The discrepancies were investigated and it turned out in a second appraisal that the control coding was the correct one. Therefore, we chose to rely on the first coding after another attempt at validating the data. Thus, in view of the less than perfect agreement between the expert rater and the first co-rater, we elicited ratings from a second co-rater after more extensive training and with clearer guidelines. For this second co-rater, we selected a different subset of stimuli with as much overlap to the first subset as possible, but at the same time with even distribution across the 3 WD dimensions (4 NON, 3 INC-RES, 3 SEX) and excluding the warm-up items 1–5 from both sets, which was not observed for the first subset. This left us with 6 stimuli overlapping between the first and the second sets. As expected, the agreement between the expert rater and the second co-rater was much better, for the qualitative data on average 76%, with the LM type, the most abstract concept, lowest at 40%, and NS, the most constrained, highest at 100%. For the quantitative data, the correlations achieved were as follows: for degree of incongruity INC (.47), degree of resolution (.36), and residual incongruity (.55). There was very little correlation between the first and the second co-raters, indicating that the instructions for the first were insufficient.

All in all, the correspondence was of a mid size and reminds us that there is some leeway even in expert coding. Obviously, these data show that caution needs to be placed on the analysis of experts. Sources of problems with the expert ratings as required for the present task are addressed in more detail in the discussion section.

6. Results

In the following, we will discuss the results of comparing the expert coding condition to the original 3 WD condition with a set of statistical tools. No special attention will be paid to the isolated results of the latter as they are discussed in detail in the sources outlined above.

While the prime emphasis is on testing differences among the three types of humor, the forms of the 3 WD should also be considered in order to see whether the two tests are indeed parallel or whether some findings are only, or to a higher degree, valid for one form but not the other. Therefore, a series of two by three ANOVAs was performed with form of 3 WD (Form A vs. Form B) and type of humor (INC-RES, NON, SEX) as independent variables and the different ratings as dependent variables.

6.1. *Is NON more incongruous than INC-RES and SEX?*

Of the 60 jokes and cartoons, 16 yielded a “1” in degree of incongruity, 26 a “2”, 11 a “3,” and 7 were highest in degree of incongruity. The 2 by 3 ANOVA for the rating of degree of incongruity yielded a main effect for type of humor ($F[2, 54] = 6.007, p = .0044$), but no effect of form or an interaction. Post hoc tests (Fishers PLSD) revealed that NON was rated significantly more incongruous than both INC-RES ($p = .0032$) and SEX ($p = .0053$), which did not differ from each other (see Figure 1).

6.2. *Is there more residual incongruity in NON than in INC-RES and SEX?*

Of the 60 jokes and cartoons, 28 yielded a “1” in degree of residual incongruity, 17 a “2”, 12 a “3,” and 3 were highest in degree of residual incongruity. The 2 by 3 ANOVA for the rating of degree of residual incongruity (rINC) yielded a main effect for type of humor ($F[2, 54] = 16.120, p < .0001$) and form of 3 WD ($F[1, 54] = 4.438, p = .0389$), but no interaction. Post hoc tests (Fishers PLSD) revealed that NON was rated to contain significantly more residual incongruity than both INC-RES ($p = .0006$) and SEX ($p < .0001$), and SEX tends to be still lower in rINC than INC-RES ($p = .0582$). The main effect of test form seems

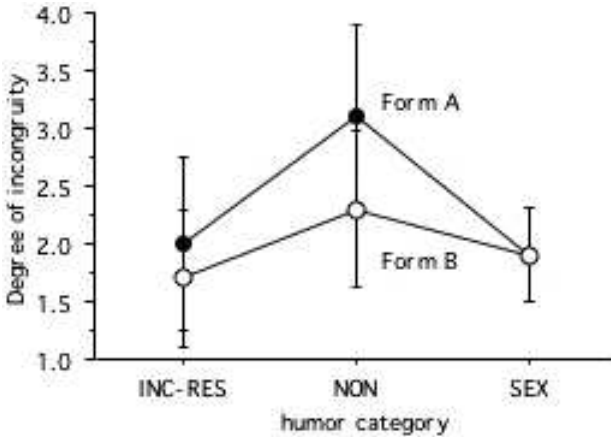


Figure 1. *Degree of incongruity as a function of humor category and form of test*

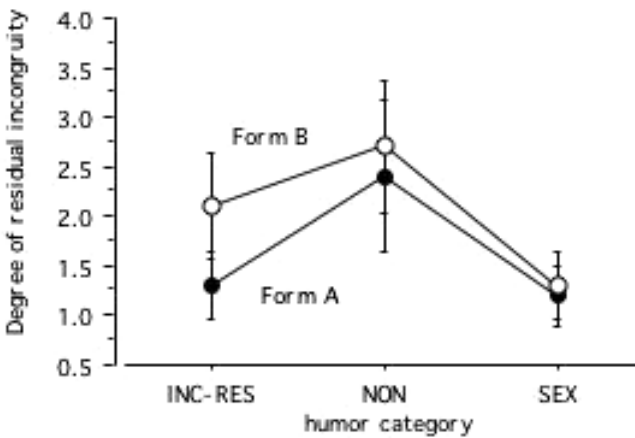


Figure 2. *Degree of residual incongruity as a function of humor category and form of test*

to be based on the fact that in form B the INC-RES items are considered to be higher in residual incongruity than SEX (see Figure 2).

6.3. *Does NON yield a lower degree of resolution than INC-RES and SEX?*

Of the 60 jokes and cartoons, 32 yielded a 1 in degree of resolution, 12 a “2”, 11 a “3,” and 5 were highest in degree of incongruity.

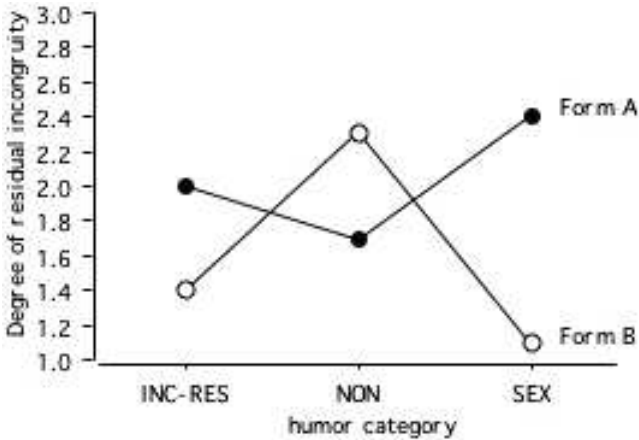


Figure 3. Degree of resolution as a function of humor category and form of test

The two by three ANOVA for the rating of degree of resolution as the dependent variable yielded no main effect for type of humor ($F[2, 54] = 0.582$, ns); i.e., overall there was no difference in degree of resolution for the three types of humor. The main effect for test form ($F[2, 54] = 3.175$, $p = .0804$) failed to reach significance, but there was a significant interaction ($F[2, 54] = 5.205$, $p = .0086$). Figure 3 shows that sexual humor in form A was rated more resolvable than in form B.

As there was a high number of jokes/cartoons that yielded the lowest score (i.e., are very simple to understand) an ANOVA might be too insensitive to reveal differences. Therefore, it is instructive to look at the contingency table. For both INC-RES and NON humor (including sexual humor) there were 12 stimuli yielding the lowest score (1 = very simple to comprehend). Eight more of the INC-RES-based items but only three NON-based cartoons were considered simple (= 2). While six INC-RES items yielded a “3” (i.e., complex), none of them yielded a very complex (= 4) to comprehend. For NON stimuli however, four were rated as complex and four more were considered very complex to understand. Thus, while there is roughly the same amount of very simple items in both, there seems to be a tendency for NON cartoons to be among the ones harder to get. However, the CHI-Square analysis failed to produce a significant result (albeit only marginally so), $\chi^2(3) = 6.531$, $p = 0.0891$.

6.4. *Do the subgroups of sexual humor differ with respect to structural properties: Comparing NON SEX, INC-RES SEX and PURE SEX*

As a content-based category, sexual jokes and cartoons were separated according to their grouping into INC-RES SEX humor (N = 6), PURE SEX humor (N = 11) and NON SEX humor (N = 3). Univariate analyses of variance were computed with type of sexual humor as classification variable and degrees of incongruity, resolution, and residual incongruity as dependent variables. There is indeed an effect of type of sexual humor ($F[2, 17] = 6.848, p = .0066$), with NON SEX being significantly more incongruous than both PURE SEX ($p = .0020$) and INC-RES SEX ($p = .0446$), which did not differ from each other. As before, there is no effect for degree of resolution ($F[2, 17] = .419, ns$), but an effect on degree of residual incongruity ($F[2, 17] = 9.793, p = .0015$). Again NON SEX was higher than both PURE SEX ($p = .0004$) and INC-RES SEX ($p = .0019$), which did not differ from each other.

6.5. *Different types of script constellations in INC-RES, NON and SEX?*

A contingency table analysis was performed for type of humor and the three major abstract types of script constellations (actual/non-actual, normal/abnormal, possible/impossible). Order of two scripts (impossible following possible vs. possible following impossible) was not considered in this analysis. Table 1 shows that actual/non-actual opposition was most frequent, followed by normal/abnormal and possible/impossible. While normal/abnormal was equally often represented in the three humor types, NON seems to be less often utilizing the actual/non-actual distinction and more often the possible/impossible SO. A CHI-square analysis of the full table did not yield significance. However, the exclusion

Table 1. *Types of script oppositions in the three factors*

Script oppositions	Humor types			Total
	INC-RES	NON	SEX	
actual/non-actual	12	7	12	31
normal/abnormal	5	6	7	18
possible/impossible	3	7	1	11
Total	20	20	20	60

of the normal/abnormal SO produces a significant result, $\chi^2(2) = 6.704$, $p = 0.0328$.

Considering the order of two scripts does not change the results much. Indeed, the “reverse” order is rather rare. “Non-actual following actual” occurred in 27 cases while “actual following non-actual” occurred only four times. Normal following abnormal did not occur at all, and of the 11 instances of involving possible/impossible 10 were impossible following possible. Thus, in 55 out of 60 cases the first script is the “real” one.

6.6. Script oppositions and degree of resolution?

No effects of SO constellation type on degree of resolution could be observed. As the rating was compatible with the proposals of Rothbart and Pien (1977), this seems to indicate that the relation between these two factors is either more complex or of an altogether different nature. Further investigation is required here.

6.7. Script oppositions and degree of residual incongruity?

A 2 by 3 ANOVA was performed with form of 3 WD (A vs. B) and type of script opposition constellations (actual/non-actual, normal/abnormal, possible/impossible).

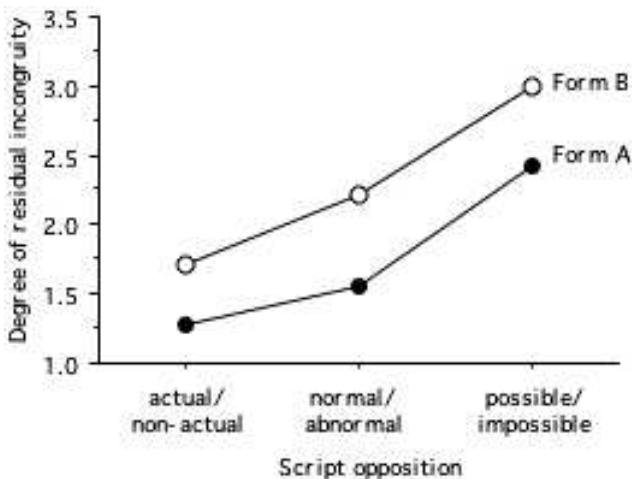


Figure 4. Degree of residual incongruity as a function of type of SO and form of test

possible/impossible) as independent variables and the rating of degree of residual incongruity as the dependent variable. There was a main effect for type of SO ($F[2, 54] = 8.619, p = .0006$), a main effect for test form ($F[1, 54] = 5.553, p = .0221$), but no interaction. Post hoc tests (Fishers PLSD) revealed that possible/impossible opposition was higher in degree of residual incongruity than both normal/abnormal ($p = .0201$) and actual/non-actual ($p = .0003$), while the latter two did not differ from each other significantly. In other words, impossible SOs yield higher residual incongruities. In addition, form B seems to have higher residual incongruities.

6.8. *Analysis of SO antonymy*

While the 3 WD has only one content category, namely sexual topics, stimuli that could be labeled “aggressive” were found in all categories. Apparently, this type of content does not seem to be strong enough to warrant a separate category. The coding included the type of script opposition antonymy. In contrast to the constellation of the two scripts that the joke text brings about, this factor aims to capture the type of oppositeness relation that holds between them. Table 2 contains the different SO antonymies identified for the three categories.

Table 2 shows that the SO antonymies of the first two factors are quite diverse. In the SEX category, the picture is very different though. Only 4 types of antonymies occurred, and the ones involving sex make up 17 of the 20 items, or even 19 if one includes love. While this is a high number, it is interesting to see that naïve subjects do respond to those 20 in a similar way (and each of them has some reference to sex in it); however, the expert coding did find reference to sex in 3 cases of the 20 INC-RES, while 3 SO antonymies in the SEX humor category seem to refer to love/non-love, and animal/human.

6.9. *Analysis of LM (Logical Mechanisms)*

Altogether 15 mechanisms were identified. They were not equal in frequency. False analogies were most frequent occurring in 22 jokes and cartoons, followed by shifts of perspectives (10) and other types of analogy (6). Others, like metaphor overapplication, ironic reversal, or being

Table 2. *SO* antonymy and the three factors

INC-RES	NON	SEX	Categories SEX
pain/cure (1)	insult/non-insult (1)	animal/human (1)	INC-RES
fame/non-fame (1)	nuisance/non-	religion/non-	religion/non-
sick/healthy (1)	nuisance (1)	religion (sex) (1)	religion (sex) (1)
clean/dirty (1)	home/non-home (1)	love/non-love (2)	sex/non-sex (5)
childhood/ technology (1)	knowledge/utility (1)	sex/non-sex (16)	NON
animal/human (1)	animal/human (1)		animal/human (1)
love/non-love (1)	violence/non-		sex/non-sex (2)
food/non-food (2)	violence (1)		PURE
life/death (2)	water/earth (1)		love/non-love (2)
feces/non-feces (2)	big/small (1)		sex/non-sex (9)
money/non-	death/life (1)		
money (4)	religion/non-		
sex/non-sex (3)	religion (1)		
	manner/non-		
	manner (1)		
	life/death (2)		
	real/unreal (2)		
	love/non-love (2)		
	feces/non-feces (3)		

Note. Numbers in brackets list the frequency of that antonymy. Different types of *SO* are ordered by frequency.

too literal, occurred only once. In order to examine whether or not some of these logical mechanisms are specific for certain humor factors, the frequencies were computed for the three humor categories separately (see Table 3).

Table 3 shows that the two structural factors utilize a large number of different logical mechanisms and none is particularly salient or specific for any of the humor factors. On the contrary, it is rather the sexual humor category that has one salient mechanism: false analogy, which occurs in 60% of the cases. An examination of the subtypes shows them to occur more often in the PURE and INC-RES subcategory of sexual humor. False analogies are also frequent in INC-RES but only make for 25% of the mechanisms here.

6.10. *Analysis of NS (Narrative Strategies)*

As the 3 WD contains both jokes and cartoons, some additional subdivisions had to be considered for the KR narrative strategy. Table 4 lists the NS groups separately for the three 3 WD factors.

Table 3. *LM and the three 3 WD humor factors*

INC-RES	NON	SEX	Categories SEX
hyperbole (1)	none (2)	pun (1)	INC-RES
being too literal (1)	revelation of prerequisite (1)	metaphor overapplication (1)	metaphor overapplication (1)
anthropomorphic role reversal (1)	figure-ground reversal (1)	shift of perspective (2)	false analogy (5)
revelation of prerequisite (2)	garden path (1)	anthropomorphic role reversal (2)	NON
pun (3)	anthropomorphic role reversal (1)	analogy (2)	anthropomorphic role reversal (1)
analogy (3)	false concept of compensation (1)	false analogy (12)	analogy (2)
shift of perspective (4)	ironic reversal (1)		PURE
false analogy (5)	purported precondition (1)		anthropomorphic role reversal (1)
	denying the obvious (1)		pun (1)
	analogy (1)		shift of perspective (2)
	shift of perspective (4)		false analogy (7)
	false analogy (5)		

Note. Numbers in brackets list the frequency of that LM. Different types of LM are ordered by frequency. “None” means that no mechanism was identified.

Table 4. *Narrative strategies involved in the three 3 WD humor categories*

Narrative strategy	INC-RES	NON	SEX
Jokes	13	0	6
narrative	11	0	6
narrative (Q-A)	1	0	0
Q-A	1	0	0
Cartoons	7	20	14
1-panel cartoon with bubble for speech	0	1	1
1-panel cartoon with caption for speech	6	3	10
1-panel cartoon with heading	1	0	1
1-panel cartoon without words	0	2	1
2-panel cartoon without words	0	3	0
2-panel cartoon, both captioned with speech	0	1	0
2-panel cartoon, one bubble	0	1	0
3-panel cartoon with speech above characters	0	1	0
3-panel cartoon without words	0	0	1
4-panel cartoon with bubbles	0	1	0
4-panel cartoon with words for dialogue in panels	0	1	0
4-panel cartoon without words	0	1	0
6-panel cartoon with bubbles	0	4	0
6-panel cartoon without words	0	1	0
Total	20	20	20

Table 4 shows that not only does NON consist exclusively of cartoons, but also of cartoons with a higher number of panels. INC-RES may be jokes or cartoons, but these cartoons typically contain only one panel. SEX humor may take the form of jokes, but a majority of it is cartoons. Nevertheless, except for one all contain only one panel.

The distribution across the subgroups of SEX humor is peculiar too. Five of the six INC-RES SEX stimuli are jokes and one is a 1-panel cartoon with a caption for speech. The NON SEX items were all cartoons with one or three panels. Again, PURE SEX items were mainly 1-panel cartoons with caption for speech (9 cases), and one 1-panel cartoon with caption for speech and one joke.

6.11. *Pornotopia*

In an independent analysis, a total of 7 items of the SEX factor were marked as containing pornotopia (cf. Dines-Levy and Smith 1988), i.e., the depiction of an exaggeratedly voluptuous female as available to an average male. They were all 1-panel cartoon with a caption for speech,

in which the LM trigger is in the text. They all obtain low residual incongruity and all except for one belong to the PURE SEX subcategory with the one exception from INC-RES SEX.

6.12. *Analysis of the KR TA (Target)*

Suls (1977) suggested that disparagement theories and his two-stage model are compatible and stereotypes about the targets may indeed help to resolve the incongruity. Nonsense might be more strongly related to mere play with ideas, and as such one might expect fewer stimuli with targets here. Of the 60 items, 26 were coded as having no target, while 34 have one. However, identifying the specific target is somewhat subjective, and in 13 of the 34 cases there were doubts about the exact target, while in 21 cases the target was clear. The distribution was not random; 60% of the NON cartoons were seen as having *no* target, while 70% and 60% of the INC-RES and SEX were seen as having one, respectively. Sexual humor may be based on either structure, and thus INC-RES SEX jokes were added to the INC-RES category, while NON SEX items were subsumed under NON. Pure SEX jokes and cartoons were ignored in the analysis. A two-way contingency table was established, which yielded a significant association (phi coefficient = .342, $\chi^2(1) = 4.465$, $p = .0346$). Indeed, of the 26 jokes and cartoons based on INC-RES, 19 (73.1%) were seen as having a target, while only 9 (39.1%) of the 23 NON cartoons had a target.

6.13. *Interrelations among coded-theoretical and derived-empirical parameters*

Product-moment correlations between the factor loadings, means in several criteria, and the coded parameters were computed and are presented in Table 5. Tests of significance are not meaningful between all variables: among single and composite scores, for example, like degree of incongruity, or among the factor loadings. Thus, tests of statistical significance were only applied where they seemed appropriate.

Table 5 shows that degree of incongruity correlates positively with degree of residual incongruity. However, surprisingly neither of them correlates with degree of resolution. Script oppositions of the possible/impossible type are conducive to more residual incongruity, and SOs of

Table 5. Intercorrelations among coded and empirical parameters

	Factor loadings			Mean rating of how			Coded degree of			Type of SO		
	INC- RES	NON	SEX	funny	aversive	known	INC	RES	rINC	actual/ non-actual	normal/ abnormal	possible/ impossible
Factor loadings												
NON	-.52											
SEX	-.09	-.66										
Mean rating of												
funniness	.34**	-.06	-.25(*)									
aversiveness	-.42***	-.29*	.67***	-.64***								
well-known	-.52***	.26*	.17	-.50***	.43***							
Coded degree of												
incongruity	-.25(*)	.50***	-.45***	-.04	-.15	.28*						
resolution	-.10	.14	-.07	.15	-.12	-.07	-.02					
residual incongruity	-.18	.54***	-.57***	.04	-.35**	.04	.51***	.04				
Type of script opposition												
actual/non-actual	.08	-.33*	.29*	.11	.17	-.20	-.57	.00	-.38**			
normal/abnormal	-.02	.07	.00	-.01	.05	.06	.13	.08	.04	-.61		
possible/impossible	-.11	.40**	-.41***	.00	-.30*	.17	.70	.00	.42***	-.44	-.31	
Target (yes, no)	.24	-.43**	.40**	.15	.04	-.24	-.33*	.11	-.39**	.27(*)	-.08	-.26(*)

Note. N = 60 (jokes and cartoons of the three categories), INC = incongruity, RES = resolution, rINC = residual incongruity.

Coefficients in italics were not tested for significance.

(*) < .05 (one-tailed); * < .05; ** < .01; *** < .001.

the actual/non-actual type are inherent in jokes and cartoons with less residual incongruity.

Both degree of initial and residual incongruity increase with the loadings on the NON factor. Prototypical NON jokes and cartoons less often have an SO of the actual/non-actual type and more frequently one of the possible/impossible category. There is no correlation with (low degree of) resolution. Having no target is more frequent for NON humor and items rate as having higher initial and residual incongruity. Jokes and cartoons that are prototypical for INC-RES tend to have incongruities of a lower degree and there is not a higher degree of resolution.

SEX yields a pattern of correlations almost opposite to that of NON. There is a lower degree of initial and residual incongruity, as well as more actual/non-actual and less possible/impossible types of script oppositions. This is partly surprising, as SEX humor should correlate less with structural properties. However, the size of the loading on the SEX factor is impaired by the presence of loadings on the structural factors. Thus, an analysis of the subcomponents of the SEX humor category as described above is, strictly speaking, appropriate.

7. Discussion

First and above all, the present study confirms that psychology and linguistics can talk to each other and conduct joint research projects successfully. The present article addresses the meta-issue of mutual confirmation for a number of smaller issues, but it does seem to be a promising first step in an interdisciplinary dialogue in humor research. Such projects do gain as their theories and methodologies complement each other to solve research questions in a manner not limited by disciplinary boundaries, but in a manner in which these issues will present themselves. The gain in the present case is mostly for humor research and psychology, while linguistics served largely as a donor. The results are summarized in Table 6.

A GTVH analysis of the 3 WD jokes and cartoons shows that no single parameter distinguishes perfectly among the three humor categories, but several of them do contribute to the characterization of the humor types arrived at by factor analysis (see Table 6). Above all, it is the degree of residual incongruity that distinguishes among the humor types. NON (and NON-based sexual humor) has the highest amount of residual

Table 6. *Summary of findings: The 3 WD categories and the original and derived GTVH-parameters*

	INC-RES	NON	SEX
Degree of incongruity	medium	high	medium (high for NON SEX)
Degree of residual incongruity	medium	high	low (high for NON SEX)
Degree of resolution	very simple to complex	very simple to very complex	-----
Script opposition	diverse	actual/not actual less often possible/impossible more often	diverse
SO antonymy	diverse	diverse	sex/non sex prevails
Logical mechanism	diverse	diverse	False analogies (especially in INC-RES and PURE)
Narrative strategy	Text, cartoons with 1 panel	Cartoons with a higher number of panels	Text, cartoons with 1 panel (NON SEX with more panels)
Pornotopia	does not apply	does not apply	prevails in PURE SEX
Target	involves targets frequently	involves targets rarely	involves targets frequently (NON SEX rarely a target)

incongruity, followed by INC-RES and SEX. Initial incongruity is a discriminating feature too, with NON exceeding INC-RES and SEX, primarily because there is a higher frequency of possible/impossible SOs and a lesser number of actual/non-actual script oppositions in NON. This is not surprising as the initial and more elaborate interpretation of nonsense included the “absurdity” element (Ruch 1980).

The analysis of script opposition antonymy showed great variety for INC-RES and NON, and only a lower number of antonymies for SEX. But even there, four different types were listed, suggesting that sex/non-sex does not exhaust the relevant varieties. Here, as elsewhere, it would be of interest to see how several trained GTVH coders converge. False analogies was the single most frequent logical mechanism, and it dominated (12 out of 20) the sexual humor category. However, false analogy was present in all humor categories and so were other logical mechanisms (e.g., analogies, anthropomorphic role reversal). Most importantly, in two cases no logical mechanism could be identified; in both cases, the stimuli had loadings for NON. This points in an interesting direction for further research on nonsense humor in relation to resolution, given that even experts have a hard time identifying the resolving LM. In these instances, the resolution must be very partial.

The KR *narrative strategy* yielded results that in part help to distinguish among the 3 WD factors. While all three categories included cartoons, only the NON category (and NON-based sexual humor) had cartoons with more than one panels. Also, in the NON category there is no joke, although in the starting material some verbal forms (e.g., limericks) did load on the nonsense factor. Obviously, more non-symbolic deviation from reality is possible with cartoons and a larger numbers of panels allow for the complexity sought after by the recipients of NON humor.

Targets were not always easily identifiable, but it was very often possible in INC-RES humor (and INC-RES-based sexual humor). Targets were often missing for cartoons operating with the NON structure. Obviously, pornotopia leads to high loadings on sexual humor and simultaneously low (or virtually zero) loadings on the structure-dominated factors.

Thus, of the six GTVH categories, it is primarily SO, NS, TA, and LM that contribute to the separation of the three humor types with respect to effects on recipients. However, degree of initial incongruity and degree of residual incongruity contributed the most. These parameters are not an integral part of the GTVH, but were operationalized on its basis. Degree

of resolution did not discriminate convincingly. Maybe the operationalization needs to be reconsidered or at least validated.

While application of the GTVH allows for a rich description of any textual corpus studied for a variety of possible research questions, this advantage was anticipated to be problematic for the present study, as the assumed pureness and non-directedness of its application here does not allow for a specification of grain size for the KR analyses. In other words, lacking a research issue other than the meta-task of mutual theoretical evaluation, there is no basis to decide on a level of abstraction in the description. This holds in particular for the content-based KRs SO, LA, and SI, but also for the structural KRs. Indeed, from the outset it turned out to be very hard to achieve interrater reliability, as pointed out above. This difficulty has led us to exclude the analysis of LA and SI for the very reason that there was no issue that could unify the analysis and rating at these levels. Sexual humor is easily identified and does not really require a GTVH interpretation. But overall, it seems to be advisable to include interrater reliability checks in linguistic analyses of humor. Also, it needs to be clear that coding requires a certain level of familiarity with a model that includes abstract notions. But similarly, it would be hard to have a non-phonologists rate tonal units or even phonemes without first training them for this task—just as it would be difficult to ask untrained music lovers to name the dominant chord and its secondary tonics in a song they just heard.

The study shows empirically that various parameters are intercorrelated, and those that are intercorrelated do discriminate among the types of humor in a similar fashion. Indeed, it seems that factor-analytic studies of jokes and cartoons have lumped humorous stimuli together mainly along a syndrome of correlated criteria. The possible/impossible script opposition is conducive to higher degrees of incongruity and residual incongruity. It emerges more often in humor that is considered aesthetically pleasing, and it does not require targets. The NON humor category combines all those features. But it has to be cautioned that these are among the most tentative results and further investigation is required. Nevertheless, NON is more precisely described by the results of the present study. Further work with a specific focus on INC-RES will be necessary, as the analysis did not yield much clarity here. With the exception of the presence of a target, INC-RES seems to be characterized by the absence or lower degree of features: a lower degree of initial and residual incongruity, and a lower number of cartoon panels.

The separation was not perfect; i.e., the three types of humor do overlap in each of the parameters. Even the degree of residual incongruity, otherwise the single best discriminating criterion, did not separate INC-RES and NON humor unequivocally. A number of different reasons may be responsible for this. Firstly, maybe a combination of parameters is required, not a single one. Secondly, as repeatedly mentioned, the coding might not be perfect. This can be seen in the interrater agreement. Obviously, the least overlap could be achieved for open or semi-open categories, such as the SO abstract or LM type. Much of the remaining lack in overlap can be ascribed to differing degrees of abstraction in the rating of open categories. Generally, a higher number of coders might have yielded more discriminative results. Thus, it seems necessary to train more researchers using this coding system and see which of the GTVH parameters can be objectively assessed.

Thirdly, while relatively pure markers of the factors were selected, they are not identical copies of each other. In particular, nonsense is known to be somewhat heterogeneous, as can be seen in the description of this factor as well as the factor loadings. Nonsense typically shows a lower Cronbach Alpha (an index of homogeneity) too. Maybe, the results can simply be taken for what they show: namely, that initial and residual incongruity distinguish between INC-RES and NON. But they are only necessary, yet not sufficient, criteria for the separation. Aesthetic quality might be a further factor, as might be cognitive complexity, but there is no operationalizable notion of either of these concepts, despite vague attempts to use these in cognitive science. Degree of resolution, possibly related to cognitive complexity, was not discriminative. This might again lie in the nature of the coding, but the problem is likely to be of a general methodological or theoretical nature: Is expert coding an appropriate measure or should fMRI or eye-tracking experiments be used?

Nonsense was rated as more incongruous, containing more residual incongruity, rarely containing a target, more often utilizing a higher number of panels and being more aesthetically pleasing than the other two categories. Those features of humor have probably developed for people with a higher need for uncertainty (in information-theoretic sense, e.g. complexity, novelty, asymmetry) in general life as well as in humor (see Ruch and Hehl 1998).

As initial and residual incongruity were the best discriminating variables (and both are not genuine elements of the GTVH), one has to ask what precisely has been gained over previous descriptions. The 3 WD

interpretations combined descriptions of the jokes' features and the process of taking up and processing of the material—all presumably through the lens of the recipient. The present study adds information to the interpretation that focuses on the material itself and its processing—not how it is perceived and appreciated by laypeople.

8. Outlook

Because the present study was based on a humor test with three *homogeneous* sets of jokes and cartoons according to the 3 WD, there might be a limited variation of these stimuli in terms of the six KRs of the GTVH. Thus, one can argue that the importance of certain KRs is underestimated and that their relative importance would appear different in a *random* set of jokes and cartoons. This is why we propose to conduct further studies with a related, yet different, methodology. For example, one might derive a random and representative set of jokes/cartoons—irrespective of any factor structure—and subject them to both a GTVH analysis and an empirical evaluation. Then a regression analysis or similar method would help to derive weights for the importance of the different KRs. Alternatively, one might search for a set of jokes/cartoons that represent all relevant variations in the KRs, in order to assure that those variations are not arbitrarily restricted by chance or sampling procedure, and then apply the same procedure. This way, one could further research the form and type of function of the relationship between KRs and perceived joke properties.

The present study provides a first instance of such an approach, based on a cooperation of linguistics and psychology (not to be confused with psycholinguistics!). That such an approach can be successful is not least witnessed by the fact that the authors still talk to each other on a very regular basis.

*Georgia Southern University
University of Zürich*

Notes

Correspondence address: hempelma@mac.com

We would like to thank Matthias Perner and Andrea C. Samson for their help in rating the stimuli, as well as the anonymous reviewers for their thoughtful comments.

1. Raskin (1985: 127) sees the following relations as frequent types of SO:

If script I is then script II is
 – actual non-actual
 – normal possible
 – possible impossible

Here, “normal/possible” is a typo for “normal/abnormal” (Raskin, personal communication). The relations are binary.

References

- Apte, Mahadev
 1988 Disciplinary boundaries in humorology: An anthropologist’s ruminations. *Humor: International Journal of Humor Research* 1 (1), 5–25.
- Attardo, Salvatore
 1997 The semantic foundations of cognitive theories of humor. *Humor: International Journal of Humor Research* 10 (4), 395–420.
- Attardo, Salvatore, and Victor Raskin
 1991 Script theory revis(it)ed: Joke similarity and joke representation model. *Humor: International Journal of Humor Research* 4 (3–4), 293–347.
- Attardo, Salvatore, Christian F. Hempelmann, and Sara Di Maio
 2002 Script oppositions and logical mechanisms: Modeling incongruities and their resolutions. *Humor: International Journal of Humor Research* 15 (1), 3–46.
- Brône, Geert, and Kurt Feysaerts
 2004 Assessing the SSTH and GTVH: A view from cognitive linguistics. *Humor: International Journal of Humor Research* 17 (4), 361–372.
- Davies, Christie
 2004 Victor Raskin on jokes. *Humor: International Journal of Humor Research* 17 (4), 373–380.
- Dines-Levy, Gail, and Gregory W. H. Smith
 1988 Representations of women and men in *Playboy* sex cartoons. In Powell, Chris, and George E. C. Patton (eds.), *Humour in Society: Resistance and Control*. New York: St. Martin’s, 234–259.
- Gernhardt, Robert
 1988 Versuch einer Annäherung an eine Feldtheorie der Komik. In *Was gibt’s denn da zu lachen? Kritik der Komiker, Kritik der Kritiker, Kritik der Komik*. Zürich: Haffmans.
- Giora, Rachel
 2002 Optimal innovation and pleasure. In Stock, O., C. Strapparava, and A. Nijholt (eds.), *The April Fools’ Day Workshop on Computational Humour*, April 2002. Trento: ITC-itst, 11–28.
- Hempelmann, Christian F.
 2004 Script opposition and logical mechanism in punning. *Humor: International Journal of Humor Research* 17 (4), 381–392.
- Mahesh, Kavi, Sergei Nirenburg, and Steven Beale
 1997a If you have it, flaunt it: Using full ontological knowledge for word sense disambiguation. In *Proceedings of TMI-97, Santa Fe, NM*, 1–9.

- Mahesh, Kavi, Sergei Nirenburg, Steven Beale, Evelyne Viegas, Victor Raskin, and Boyan Onyshkevych
1997b Word sense disambiguation: Why statistics if we have these numbers. In *Proceedings of TMI-97, Santa Fe, NM*, 151–159.
- McGhee, Paul E., Willibald Ruch, and Franz-Josef Hehl
1990 A personality-based model of humor development during adulthood. *Humor: International Journal of Humor Research* 3, 119–146.
- Nirenburg, Sergei, and Victor Raskin
2004 *Ontological Semantics*. Cambridge, MA: MIT Press.
- Onyshkevych, Boyan
1997 An ontological-semantic framework for text analysis. Unpublished Ph.D. thesis, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, PA.
- Paolillo, John C.
1998 Gary Larson's *Far Side*: Nonsense? Nonsense! *Humor: International Journal of Humor Research* 11 (3), 261–190.
- Raskin, Victor
1985 *Semantic Mechanisms of Humor*. Dordrecht: D. Reidel.
1995 The state of the art in humour research: Let's amalgamate! In Paton, George E. C., and Robert W. Jones (eds.), *13th International Humor Conference*. Birmingham, UK: University of Aston, 44.
- Raskin, Victor, Sergei Nirenburg, Christian F. Hempelmann, Inna Nirenburg, and Katrina E. Triesenberg
2003 The genesis of a script for bankruptcy in ontological semantics. In *Proceedings of the Text Meaning Workshop of the HLT-NAACL 2003 Human Language Technology Conference*, Edmonton, Canada, May 31. Available at: <http://acl.ldc.upenn.edu/W/W03/W03-0905.pdf>.
- Rothbart, Mary K., and Diana Pien
1977 Elephants and marshmallows: A theoretical synthesis of incongruity-resolution and arousal theories of humour. In Chapman, Antony J., and Hugh C. Foot (eds.), *It's a Funny Thing, Humour*. Oxford: Pergamon, 37–40.
- Ruch, Willibald
1980 Gemeinsame Strukturen in Witzbeurteilung und Persönlichkeit. Versuch einer empirischen Integration des Gegenstandsbereiches Witzbeurteilung in die differentielle Psychologie mittels dreimodaler Faktorenanalyse. Unpublished Ph.D. dissertation, University of Graz, Austria
1992 Assessment of appreciation of humor: Studies with the 3 WD humor test. In Butcher, J. N., and C. D. Spielberger (eds.), *Advances in Personality Assessment*, vol. 9. Hillsdale, NJ: Erlbaum, 27–75.
1995 Humor-Test 3 WD. Unpublished manuscript, University of Düsseldorf, Germany.
- Ruch, Willibald, Salvatore Attardo, and Victor Raskin
1993 Towards an empirical verification of the general theory of verbal humor. *Humor: International Journal of Humor Research* 6, 123–136.
- Ruch, Willibald, and Franz-Josef Hehl
1985 Diagnose des Humors—Humor als Diagnostikum. In Hehl, Franz-Josef, Volker Ebel, and Willibald Ruch (eds.), *Diagnostik psychischer und psychophysiologischer Störungen*. Bonn: Deutscher Psychologen Verlag, 253–325.

- 1987 Personal values as facilitating and inhibiting factors in the appreciation of humor content. *Journal of Social Behavior and Personality* 2, 453–472.
- 1998 A two-mode model of humor appreciation: Its relation to aesthetic appreciation and simplicity-complexity of personality. In Ruch, Willibald (ed.), *The Sense of Humor: Explorations of a Personality Characteristic*. Berlin/New York: Mouton de Gruyter, 109–142.
- Samson, Andrea C., and Christian F. Hempelmann
n.d. Cartoons: Cognition and aesthetics. In preparation.
- Suls, Jerry M.
1977 Cognitive and disparagement theories of humour: A theoretical and empirical synthesis. In Chapman, Antony J., and Hugh C. Foot (eds.), *It's a Funny Thing, Humour*. Oxford: Pergamon, 41–45.

