

What's so funny? Modelling incongruity in humour production

Rachel Hull, Sümeyra Tosun & Jyotsna Vaid

To cite this article: Rachel Hull, Sümeyra Tosun & Jyotsna Vaid (2016): What's so funny? Modelling incongruity in humour production, *Cognition and Emotion*

To link to this article: <http://dx.doi.org/10.1080/02699931.2015.1129314>



Published online: 22 Jan 2016.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

What's so funny? Modelling incongruity in humour production[†]

Rachel Hull^a, Sümeyra Tosun^b and Jyotsna Vaid^a

^aDepartment of Psychology, Texas A&M University, College Station, TX, USA; ^bDepartment of Psychology, Süleyman Şah University, Istanbul, Turkey

ABSTRACT

Finding something humorous is intrinsically rewarding and may facilitate emotion regulation, but what creates humour has been underexplored. The present experimental study examined humour generated under controlled conditions with varying social, affective, and cognitive factors. Participants listed five ways in which a set of concept pairs (e.g. MONEY and CHOCOLATE) were similar or different in either a funny way (intentional humour elicitation) or a “catchy” way (incidental humour elicitation). Results showed that more funny responses were produced under the incidental condition, and particularly more for affectively charged than neutral concepts, for semantically unrelated than related concepts, and for responses highlighting differences rather than similarities between concepts. Further analyses revealed that funny responses showed a relative divergence in output dominance of the properties typically associated with each concept in the pair (that is, funny responses frequently highlighted a property high in output dominance for one concept but simultaneously low in output dominance for the other concept); by contrast, responses judged not funny did not show this pattern. These findings reinforce the centrality of incongruity resolution as a key cognitive ingredient for some pleasurable emotional elements arising from humour and demonstrate how it may operate within the context of humour generation.

ARTICLE HISTORY

Received 12 February 2015
Revised 21 October 2015
Accepted 2 December 2015

KEYWORDS

Humour production;
incongruity; creativity;
emotion regulation;
structural alignment

Humour is a form of mental play that has cognitive, emotional, social, and performance aspects. The capacity to perceive, produce, and enjoy humour appears to be present in all known societies, past and present. Questions about the nature, origins, and significance of humour have long engaged scholars in philosophy, anthropology (e.g. Douglas, 1975/1991), linguistics (e.g. Aarons, 2012; Chafe, 2007), psychology (Schmidt, 2002; Vaid, 2006; Veale, 2008; Wyer & Collins, 1992), and neuroscience (Samson, Hempelmann, Huber, & Zysset, 2009; Vaid & Kobler, 2000) and have led to debates about aesthetic, social, computational, cognitive, neurocognitive, and even evolutionary aspects of humour and laughter.

Attempts to theorise about humour are complicated by the sheer variety of forms (e.g. jokes, riddles, puns, stories), techniques (exaggeration, mimicry, ambiguity, sarcasm, etc.), and types of humorous

discourse (aggressive, sexual, sexist, absurd) as well as by social and contextual factors. For example, certain topics may be considered taboo and these may vary across cultures; there are also societal and/or cultural norms for when a non-serious mode of discourse is acceptable and when it is not. Sociopragmatic factors (e.g. who is the humour initiator, the audience, or the target) may influence the “fate” of a humorous utterance, that is, whether it comes across as wildly funny or mildly amusing or even as highly offensive. Given that humour often unsettles established beliefs or practices, the extent to which individuals to whom the humour is directed are “affectively committed” to those beliefs (see Veatch, 1998) may affect their judgments of and their response to utterances intended to be humorous.

Despite these challenges and complexities in understanding humour, several theories of humour

[†]Experiment 1 was presented at the Northeast Texas Humor Research Conference, Texas A&M University—Commerce, TX, October 2011.

CONTACT Rachel Hull  queenofpsyc@gmail.com

have been put forth. One prominent theory, attributed to Freud (1905/1960), focuses on the form and content of jokes, noting that many jokes refer to topics that are transgressive or taboo in some way. Freud argued that censors in the mind repress “forbidden” (e.g. hostile or sexual) thoughts and jokes represent a way for these censors to be bypassed. The form of jokes (i.e. their double meanings) is not accidental in this view, as jokes are designed to trick the censors. A cognitive analogue of this view was proposed by Minsky (1984) who suggested that humour serves to monitor and suppress “bugs” or ineffective thought processes in our everyday reasoning. This view of humour can account for a broader range of humour, including so-called nonsense humour, than that covered by the psychoanalytic account. Besides these proposed explanations, other influential conceptual frameworks have emphasised social or political aspects of humour, such as its role in exploiting stereotypes of groups, and its role in righting wrongs, such as by providing a way for socially oppressed groups to express their perspective (see Vaid, 1999, for further discussion).

Although differing in their focus, most theories of humour would concur that experiencing humour is intrinsically rewarding and can lead to feelings ranging from satisfaction to comfort or relief. Given this positive emotion-inducing aspect of humour, a number of studies have begun to explore whether and how humour may mitigate either the current experience, or later memory, of negative emotion-eliciting events. According to one view, humour provides relief by offering cognitive distraction from negative stimuli (see Strick, Holland, Van Baaren, & Knippenberg, 2010). Another view holds that humour serves to regulate negative emotion through both short-term and longer-term cognitive reappraisal of the adverse event (see Samson, Glassco, Lee, & Gross, 2014). A recent study supports the cognitive reappraisal account (Kugler & Kuhbandner, 2015) as it found that humorous reappraisal (more than rational reappraisal) counteracted negative emotions elicited by adverse experiences both in the present and in subsequent recall of those experiences, while not impairing recognition memory of the events, which would not be predicted by a cognitive distraction view (according to which both recall and recognition would be impaired).

Another line of psychological research on humour has focused on the assessment of, and individual differences in, humour appreciation (see Hempelmann &

Ruch, 2005; Ruch, 1992). One potential source of difference that has been extensively studied is gender. It appears that men and women do not differ in the frequency of perceiving, engaging in, or enjoying conversational humour (see Martin, 2014, for a review). Where gender differences are noted, they largely appear in the uses of conversational humour: women appear to use humour to create solidarity by sharing funny stories of what happened to them, whereas men appear to use humour to assert dominance (Coates, 2014). Further, both men and women appear to use humour somewhat differently depending on whether they are in same-gender or mixed-gender settings (Martin, 2014). Moreover, studies examining differences in types of humour preferred show that, where differences occur, they are in the direction of a preference for aggressive, hostile, or dark humour by men and a preference for nonsense humour more by women (Aillaud & Piolat, 2012). Men also prefer canned or formulaic humour more (Martin, 2014).

An evolutionary perspective on gender differences in humour initiation claims that humour may have evolved to signal intelligence in a prospective mate, and thus that males are likely to advertise it as an attractive trait in mate selection (Gueguen, 2010; Miller, 2000). Indeed, studies of personal ads show that men are more likely to advertise an ability to make others laugh whereas women are more likely to indicate wanting someone to make them laugh (Martin, 2014). Some experimental studies of humour production “on demand” similarly suggest that men initiate humour more often than women and that their humour is judged to be funnier (Greengross & Miller, 2011). However, other studies have found no gender differences in the incidence or success of humour production and it has been noted that even where differences may be observed there could be other ways of interpreting them (e.g. different patterns of socialisation) besides an evolutionary account (see Martin, 2014, for further discussion). Whatever the status of the evolutionary account, there is clearly a need for more research that examines actual instances of humour production, as distinct from studies of the desirability of initiators of humour. This was one aim of the present research.

Another aim was to examine the cognitive underpinnings of humour. Viewed cognitively, humour is first and foremost a creative act. Koestler (1964) proposed that humour involves *bisociative* processing; that is, processing along “two self-consistent but habitually incompatible frames of reference” (p. 35).

The element of incompatibility is a recurrent theme in most cognitive approaches to humour, which characterise humour as the byproduct of bringing into juxtaposition concepts, schemata, or meanings that are incongruous and normally would not co-occur (see Attardo & Raskin, 1991; Coulson, 2001).

The notion that humour involves a perception of incongruity and its resolution was formalised in a three stage information processing model by Suls (1972; see also Attardo, 1997; Hillson & Martin, 1994). According to Suls's model, humour perception involves three stages: a set up stage where an expectation is created, an incongruity stage where the expectation is violated, and a resolution stage where the discrepancy between the expected and the actual state of affairs is bridged. For some kinds of humour there is no resolution of the incongruity (e.g. nonsense humour). Whether the expected meaning and the actual (joke) meaning are activated simultaneously or successively has been the subject of debate. Suls's model favours a temporal separation of the various stages of meaning activation (which is endorsed by scholars such as Giora, 1991) but other scholars (e.g. Attardo, 1997; Coulson, Urbach, & Kutas, 2006; Norrick, 1986; Veatch, 1998) hold that both surprising and expected meanings must be maintained at least long enough for the perceiver to recognise that there is an incongruity. A time course study of joke meaning activation using one liner jokes provided support for the concurrent activation view but also found that the initially favoured meaning dissipates relatively quickly (Vaid, Hull, Heredia, Gerken, & Martinez, 2003). It may be that for certain kinds of humour (e.g. puns or irony) the two meanings have to be concurrently activated whereas for other kinds of humour the joke meaning erases the initially favoured meaning.

Most studies of cognitive mechanisms underlying humour have focused on humour perception (e.g. Canestrani & Bianchi, 2012; Cunningham & Derks, 2005; Du et al., 2013; Kana & Wadsworth, 2012; Samson, 2012). Less attention has been paid to mechanisms underlying humour production (but see Kozbelt & Nishioka, 2010; Mickes, Walker, Parris, Mankoff, & Christenfeld, 2012). It is important to study humour production because productive mechanisms are central to the development of cognitive theory inasmuch as productivity generates creativity, novelty, and modal symbol systems or schemata (see Barsalou & Prinz, 1997). Relying only on perception-oriented paradigms may constrain creativity by

underspecifying the range of retrieval processes involved (Blanchette & Dunbar, 2000). Relatedly, to understand cognitive mechanisms underlying humour generation it is important to study actual instances of humour generated by research participants in the context of a controlled study.

Present research

The present research sought to examine the role of incongruity in humour production by developing a controlled way of eliciting humour in a laboratory setting. Building on previous theorists' insights that unexpected interpretations of meaning often give rise to humour, the task we developed to study humour generation had participants activate and compare or contrast the meanings of pairs of discrete concepts. In selecting this way of studying humour we are not claiming that concept comparison is representative of what goes on in all types of humour, only that we believe that much humour arises when there is a play between a dominant and an uncommon response to a situation and that the conceptual comparison task allows one to capture this key element of humour. Further, we recognise that our particular way of eliciting humour foregrounds cognitive aspects more so than social aspects that would be likely to influence actual humour generation in conversational contexts. Our study nevertheless manipulated a number of variables. These included task (participants were either to find differences between the concepts or find similarities), semantic relatedness (some of the concepts were related in meaning, others were unrelated), affective dimensions of the stimuli (specifically, whether one of the concepts in a pair involves a taboo topic, or something that is normally considered socially inappropriate to talk about) and the humour producers' gender. Finally, given prior work in creativity studies which found that participants explicitly instructed to be creative were inhibited by such instructions (Runco, Illies, & Reiter-Palmon, 2005; but see Niu & Sternberg, 2001) we wondered whether a similar inhibiting effect might arise for humour elicitation. We thus compared humour elicitation when it was an incidental versus an explicit aspect of the task, by asking participants to be catchy in their responses or to be funny.

Hypotheses

We reasoned that it would be more difficult (constraining) for individuals to come up with humorous responses when explicitly instructed to be funny than

when not explicitly instructed. That is, when participants are instructed simply to be catchy we expected they would feel freer to produce humorous responses.

With respect to gender, our study allows a test of the fitness-based claim that men are more likely to initiate humour than women, and does so using a task that is different from that used in previous studies of humour production in a laboratory setting, for example, cartoon caption generation.

With respect to claims about the motivational underpinnings of humour as a defence against normally repressed thoughts (Freud, 1905/1960), we predicted that in our study concept pairs containing a taboo topic (e.g. sex, cheating) would be more likely to generate humorous responses than those containing items that were not socially taboo.

Finally, with respect to cognitive factors, in particular, incongruity, we predicted that comparisons between concepts that are unrelated in meaning would generate more incongruous responses and thus, be more likely to be perceived as humorous than comparisons between related concepts, and that focusing on how concepts differ would lead to more incongruity (and humour) than identifying how they are similar. Another way in which incongruity may operate in this task is through the particular ways in which properties shared by the concepts are brought into alignment. To examine this, we conducted a follow-up study (Experiment 2) in which we examined properties associated with the concepts in the previous experiment, as generated by another set of participants, to get a measure of their relative output dominance. We hypothesised that concept comparisons that make use of a property that is high in output dominance for one concept but lower in output dominance with the other concept in the stimulus pair will be more often judged funnier than responses that identify properties that are equivalent in their relative output dominance for the two concepts of a pair.

Experiment 1: factors influencing humorousness of elicited concept comparisons

As already noted, a humour generation task was developed in which a set of related and unrelated pairs of concepts were presented to participants who were instructed to think of how each member of a pair was either similar or different from the other member. The comparisons were in turn to be expressed in either a funny way or in a catchy way.

Method

Phase 1: humour elicitation via concept comparison

Participants and procedure

Participants were 323 college undergraduates (155 males, 168 females) from a large southwestern university who received experimental credit for participating in the study. They were randomly assigned to one of two task conditions (finding similarities, $n = 181$ vs. finding differences, $n = 142$) and to 1 of 2 instructional conditions—incidental humour production (“be catchy”), $n = 117$ or intentional humour production (“be funny”), $n = 206$. See Table 1. The concept pairs in the two instructional sets were identical, and the pair presentation order was counterbalanced.

Half of the concept pairs had been previously determined to be relatively similar in meaning (related), whereas the remaining pairs were dissimilar (unrelated). Each participant received an equal number of related and unrelated concept pairs. Further, half of the concept pairs contained an affectively charged concept whereas the remainder did not. About half of the participants ($n = 151$) were given only non-taboo concept pairs whereas the remaining participants ($n = 172$) received pairs that included one taboo¹ concept per pair.

The design was a $2 \times 2 \times 2 \times 2 \times 2$ mixed factorial. The between subjects variables were instructional set (intentional vs. incidental), participants’ gender (female vs. male), affective salience (presence or absence of a taboo item in the concept pair), and task (find similarities vs. differences). The within subjects variable was concept relatedness (similar or dissimilar).

Phase 2: classification of Phase 1 responses as funny or not funny

Participants and procedure

A panel of judges consisting of four undergraduate research assistants individually reviewed the elicited responses on the concept comparison task (Phase 1) and classified them as Funny or Not Funny, marking their judgments on a response sheet. Since the aim

Table 1. Concept pairs used in Experiment 1, Phase 1.

Relatedness	Affectively salient	Affectively neutral
Related	Love and sex Pranking and cheating	Love and friendship Pranking and joking
Unrelated	Money and sex Cooking and cheating	Money and chocolate Cooking and joking

was to get a consensus judgment, the judges were asked to mark a response as “funny” if, in their opinion, it could be considered funny at all in principle, even if the particular humour did not appeal to the judges themselves. All other responses were classified as “not funny”. Judges were not told the gender of the response producer or whether the response was elicited under intentional or incidental humour production conditions.

Two other undergraduate research assistant coders, naïve to the predictions of this study, tallied the “funny”/“not funny” judgments. Only responses on which there was consensus among three of the four judges for “funny” responses were included in the data analysis reported below.

Data analysis

The dependent variable was perceived humorousness, expressed as the mean percentage of responses judged “funny” per concept pair per participant (from a maximum of five). This was examined in a five-way analysis of variance as a function of participant gender (that is, whether the responses were generated by men or women), affective salience (presence or absence of a taboo item), instructional set (to be funny vs. catchy in expressing how the concepts of a pair were similar or different), concept relatedness (whether the concepts per pair were semantically related or unrelated), and task (to find similarities vs. to find differences between the concepts of a pair), with repeated measures on relatedness.

Results

As to be expected, there was some variability in the incidence of humorous responses generated for the different stimuli. Nevertheless, for over two thirds of the responses, one or more of the five comparisons of each concept pair produced by participants were judged to be funny by three of the four judges. Further, there was relatively high consensus across the judges in classification of responses as funny or not funny.

Sample responses

Examples of actual responses of participants are provided below, collapsed across task and instructional set. They are listed without regard to how they were classified (funny or not funny). The gender of the respondent is given in parentheses.

A. Unrelated concepts, no taboo item: MONEY AND CHOCOLATE

- one swells the wallet, the other swells the hips (F)
- neither lasts very long (M)
- both run out quickly (F)
- both smell good (M)
- both make hands dirty if you hold them too long (F)
- too much of each can be bad (M, F)
- not enough of each is not good (F)

B. Unrelated concepts, including taboo item: MONEY AND SEX

- you can't get enough (F)
- both cause problems (M)
- men want sex for money, and women want money for sex (M)
- both get people in trouble (M)
- you use one to get the other (F)
- both are hard to come by (M)
- people lie about having it/doing it (F)
- you can go to the bank for both (F)

C. Related concepts, no taboo item: LOVE AND FRIENDSHIP

- you have to be willing to compromise in both (M)
- both have the letter “e” (M, F)
- both can cause stress and anxiety (M, F)

D. Related concepts, including taboo item: LOVE AND SEX

- I don't think they're at all alike (F)
- Everyone loves to have sex (M)

Findings from analysis of variance

A five-way analysis of variance of mean percent of responses (out of a maximum of five possible responses per concept pair) that were judged to be humorous by consensus of the judges revealed three main effects and several two-way interactions.

Humour elicitation condition

A main effect of instructional set, [$F(1, 307) = 23.08, p < .001, \eta_p^2 = 0.07$] indicated, as expected, a higher incidence of funny responses when participants were instructed to “be catchy” than when they were explicitly

instructed to “be funny” (12.43% and 6.92%, respectively). Instructional set interacted with all of the other variables.

Gender

Although there was no main effect of gender an interaction of gender with instructional set [$F(1, 307) = 5.8, p < .05, \eta_p^2 = 0.02$] showed the following: both men and women showed a higher incidence of humour in the incidental than in the intentional humour elicitation condition but this effect was more pronounced for men. Men’s mean scores were 13.6% and 5.66% for catchy vs. funny responses, respectively, $t(153) = 4.84, p < .001, p' = .004^1$; women’s mean scores for catchy vs. funny response conditions were 11.23% and 8.02%, respectively; $t(166) = 2.01, p = .046, p' = .13$. When instructed to produce funny responses women showed a marginally higher incidence of humorous responses than men; $t(204) = 2.01, p = .045, p' = .13$. When instructed to be catchy there was no significant difference between men’s and women’s funny responses ($t(115) = 1.06, p = .29, p' = .29$). See Figure 1. In other words, men in the “be catchy” condition were funnier than men in the “be funny” condition, but women were just as funny as men in the “be catchy” condition, and showed a marginal tendency to be funnier than men in the “be funny” condition.

Affective salience

There was no main effect of affective salience but the interaction of this variable with instructional set approached significance [$F(1, 307) = 3.4, p = .066$].



Figure 1. Mean percent of responses judged funny in Experiment 1 as a function of humour producers’ gender and instructions to be funny vs. be catchy.

Inspection of the means indicated a higher incidence of humour when a taboo item was present in the concepts to be compared than when it was not present, and this pattern was more evident under incidental humour elicitation conditions [6.13% vs. 13.14% for taboo word pairs with to be funny and to be catchy instructions respectively; 7.84% vs. 11.64% for neutral word pairs with to be funny instruction and to be catchy instructions respectively].

Semantic relatedness

A main effect of relatedness [$F(1, 307) = 44.04, p < .001, \eta_p^2 = 0.16$] indicated a higher incidence of funny responses when the concepts to be compared were dissimilar than similar in meaning. An interaction of relatedness and instructional condition [$F(1, 307) = 4.53, p < .05, \eta_p^2 = 0.02$] revealed that the effect of relatedness was particularly strong in the incidental humour elicitation condition (15.7% vs. 9.13%, for unrelated vs. related pairs, respectively), [$t(116) = 5.17, p < .001, p' = .004$]; in the intentional humour elicitation condition the respective means were 8.72% for unrelated pairs and 5.11% for related pairs [$t(205) = 4.38, p < .001, p' = .004$]. See Figure 2.

Task—find differences vs. similarities

A main effect of task [$F(1, 307) = 39.92, p < .001, \eta_p^2 = 0.12$] indicated a higher incidence of funny responses when differences rather than similarities between the concepts were to be identified (12.48% vs. 6.12%, respectively). An interaction of task and instructional

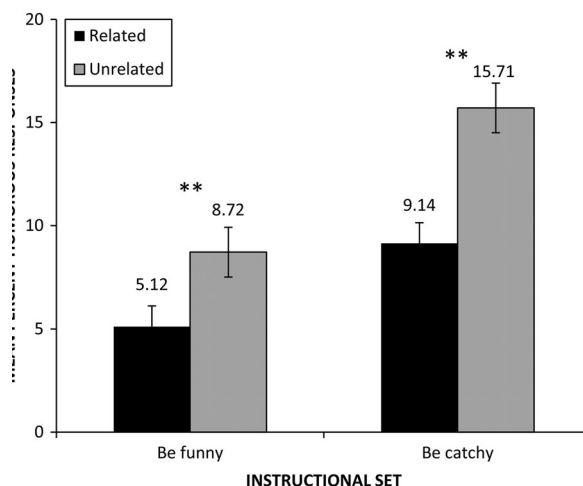


Figure 2 Mean percent of responses judged funny in Experiment 1 as a function of related vs. unrelated concepts and instructions to be funny vs. be catchy.

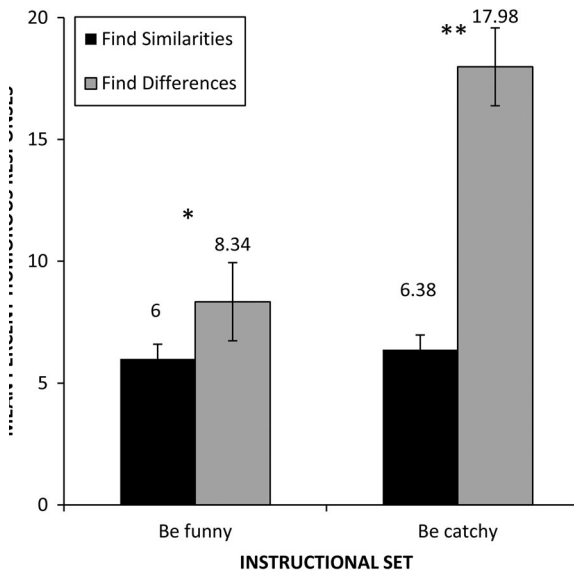


Figure 3. Mean percent of responses judged funny in Experiment 1 as a function of finding similarities vs. differences between concepts and instructions to be funny vs. be catchy.

set [$F(1, 307) = 20.42, p < .001, \eta_p^2 = 0.06$] further showed that more humorous responses were made in the “be catchy” condition than in the “be funny” condition but only for participants assigned to find differences between the concept pairs (mean scores were 17.98% and 8.34% for catchy vs. funny responses, respectively), [$t(140) = 4.91, p < .001, p' = .004$]. For participants asked to find similarities between the concepts performance was uniformly low regardless of instructional set [$t(179) = .33, p = .74, p' = .74$]; mean scores were 6.38% and 6% respectively for the incidental vs. intentional conditions. Furthermore, in the “be funny” condition, there was only a marginal effect of task [$t(204) = -1.95, p = .052, p' = .10$], whereas in the “be catchy” condition, humour production was higher on the “find differences” than “find similarities” task [$t(115) = -5.87, p < .001, p' = .004$]. See Figure 3.

Discussion

Experiment 1 explored contextual, social, affective, and cognitive factors affecting the relative incidence of funny concept comparisons responses generated by participants.

Previous research has suggested that creativity is stifled under conditions when there is pressure to be creative (Runco et al., 2005). We wondered if the

same would be found for humorous creativity. Our findings strongly indicate this to be the case: responses that were judged funny occurred more often under conditions where the humour was produced incidentally. That is, participants produced humorous responses more often when instructed to make their responses catchy rather than when instructed to make them funny. The finding that being instructed to be funny works against actually being funny underscores the creative nature of humour production. Although one might have expected that being instructed to be catchy (also a creative task) might be inhibiting as well, our findings suggest that this was not the case. In fact, humour was used as a way of being catchy. It is possible that had we configured our humour generation condition somewhat differently, for example, asking participants to come up with a humorous remark directed at a friend, we may have found a higher incidence of humorous responses. That is, the task of making comparisons between concepts might have been perceived as artificial and thus more challenging. By contrast, the directive to “be catchy” in the context of making comparisons between concepts might have been perceived as less artificial. Moreover, being asked to be catchy would have directed participants to think of aesthetically pleasing ways of expressing the concept comparisons. Other studies of aesthetic production show that humour is not uncommon when expressing something in a pleasing way, as humour is aesthetically pleasing and involves novelty and surprise (see the optimal innovation hypothesis by Giora et al., 2004 and Vaid, 2014).

Whatever the underlying reason for its effect, the variable of incidental vs. intentional humour elicitation in the present study interacted with almost all the other variables manipulated in the study, suggesting that factors that are likely to influence humour generation are more likely to do so when the humour is generated incidentally. We turn next to these other factors.

With respect to the variable of gender, fitness-based claims among some scholars of humour have led to the prediction that there should be a higher incidence of humour generation by men than by women. Our study, however, found no support for such an effect. As such, our findings do not concur with those of Greengross and Miller (2011) who found that on a cartoon caption generation task, humorous captions generated by men were judged funnier than those generated by women. Our study

found no effect of gender under incidental humour generation conditions and there was a trend towards *women* generating more humour than men under intentional humour generation conditions. Thus, at least on a laboratory presented humour production task of the sort developed in the present study no support was found for the notion that men are better at producing humour. Given that gender differences in humour production have been more often observed in the uses of humour in conversational contexts (Coates, 2014), or in the desirability of potential mates, the lack of gender differences in humour performance on an elicited humour generation task performed in a laboratory setting without targeted interlocutors of a particular gender may not be altogether surprising. At the very least, our findings may be taken to indicate that the cognitive capacity that underlies the ability to generate humorous comparisons between concepts is not differentially manifest in women and men.

With respect to the role of the affective content of the humour produced, we had hypothesised that more funny responses would be elicited for pairs containing affectively charged (“taboo”) items than for those not containing such items. Our findings did not support this hypothesis statistically although the results were in the expected direction. The modest nature of this effect in our study could be due to a lack of power, given that only a subset of the stimuli contained taboo items. Had we used a larger number of such items and perhaps a broader array of affectively charged concepts (i.e. not just pertaining to sex, but other affect-laden topics such as death, terrorism, natural disasters, etc.) a stronger effect of affective salience may have emerged. This may be an avenue for future investigation.

Finally, our study examined the cognitive basis of humour by manipulating incongruity in two ways, by varying the semantic relatedness of the stimuli, and by varying the nature of the task demands. As hypothesised, humour was produced more often when the task was to identify differences among concepts (independently of whether the concepts were similar or different), and also when the concepts to be compared were dissimilar (independently of whether the task was to find similarities or find differences). Moreover, the “humour effect” in both cases was heightened when humour elicitation occurred under incidental conditions, as already noted.

Having established that humour generation is more prevalent when the stimuli and/or task

emphasises incongruity (e.g. finding differences between unrelated concepts), the next step in our research strategy was to determine whether funny responses produced under such conditions could be predicted based on a fine-grained and quantifiable measure of incongruity. This led to our second experiment.

Experiment 2

How incongruity in structurally aligned properties predicts humour content in concept comparisons

This experiment sought to establish whether funniness can be predicted in terms of the relative output dominance values for the property a response activates to *align* two concepts (to draw a similarity or difference between them). (Output dominance is assessed by examining the relative frequency of mention of different properties associated with a concept.) Humour may be more likely to occur when there is more cognitive distance between *favoured* (output dominant) and *surprising* (unexpected) interpretations of that aligning feature, such that the feature’s default output dominance score will be high for one concept and low for the other. This is because properties high in output dominance function as default fillers and presumably reflect the most commonly activated mental representation of a concept—and the one that would therefore be initially activated during the setup stage of humour processing (i.e. before the mental reorganisation required to understand the punch line). For example, the feature *laugh* is high in output dominance for the concept *joking*, but significantly lower for the concept *cooking*, so the default output dominance scores for *laugh* would be considered misaligned for the pair COOKING and JOKING. We expected the mental reorganisation of understanding a punch line to involve the sudden alignment (equalisation of dominance) of previously misaligned properties, resulting in that singular “aha moment”—and funniness.

Our working hypothesis was as follows:

- (a) The alignment property activated by a verbal production judged Funny will be mentioned significantly more often in the base concept (high output dominance for the feature) than in the target concept (low output dominance for the feature), and;

- (b) There will be no significant difference in output dominance levels for the activated alignment property for items judged Not Funny.

To empirically test this hypothesis, we evaluated a subset of Funny and Not Funny responses from the concept comparison task in Experiment 1. The alignment property for each response was matched to the features generated for each parent concept (generated in Phase 2 of Experiment 2) to determine relative output dominance.

Method

Phase 1: stimulus development and procedure

Equal numbers of a subset of responses that had been judged Funny or Not Funny in Experiment 1 were used as stimuli in Experiment 2. The dataset contained a total of 400 randomly chosen responses from Experiment 1, including 100 from each of the 4 conditions: incidental-taboo, incidental-non-taboo, intentional-taboo, and intentional-non-taboo (where the taboo pairs were MONEY and SEX and COOKING and CHEATING and the non-taboo pairs were MONEY and CHOCOLATE and COOKING and JOKING). Importantly, we only used responses associated with unrelated concept pairs and for which participants generated responses about how the pairs were different (refer to Table 1), as Experiment 1 outcomes showed these to be significantly more conducive to humour production than comparing related pairs and finding similarities between them.

The productions were coded as Funny or Not Funny and alignment properties were then extracted from responses to each concept pair (e.g. the alignment property "hot water" was extracted from the Funny response "cooking and cheating can both get you in hot water"). Finally these alignment properties were matched to features generated from the concepts (see below).

The corpus of 400 productions was subdivided into 16 cells where each cell contained productions from a single concept pair which were to be used as stimuli for participants in Phase 2 of this experiment (described below). Specifically, the 400 responses were first divided into their parent instructional condition categories of incidental and intentional humour. Within each condition category, responses were further subdivided into humour groups of Funny and Not Funny (based on judges'

consensus). Finally, within each of these response types responses were sorted into four cells, each representing the particular concept pair from which responses were generated (i.e. money and chocolate, money and sex, cooking and cheating, and cooking and joking).

The upshot of this subdivision scheme was that each participant in Experiment 2, Phase 2 (described below) would receive no more than 5 stimuli drawn from a single concept pair (e.g. 5 Funny responses to the concept pair "cooking and joking" under the intentional humour condition), but actual responses varied by concept pairs for two reasons. First, some pairs elicited higher numbers of unique funny responses than other pairs, and second, Not Funny responses exceeded Funny responses in general. To equalise the number of Funny and Not Funny responses across the 16 cells, we randomly dropped excess Not Funny responses, resulting in a final total of 29 Funny and 29 Not Funny responses.

Identification of alignment properties

Two undergraduate coders examined each of the 58 responses (29 Funny and 29 Not Funny) taken from

Table 2. Alignment property stimulus sets drawn from responses generated in Experiment 1.

Money and chocolate	Money and sex	Cooking and cheating	Cooking and joking
Funny productions			
negative outcome	think about	shortcuts	makes you sick
leaves hands quickly	politician	denial	causes discomfort
stress relief/escape	lack of	hurt	laugh
fattening	depletion	spicy	right ingredients
evil	awkward	meat	chicken
sweet	hard to get	not fatal	dad/male
melt	diminish	hot water	
	never had	clean up after	
Not Funny productions			
satisfying	bad	practice	skill/practice
wanted	dirty	skill	produce enjoyment
easy to get	evil	bad outcome	spicy
touchable	highly desired	give you gain	wait
attractive to women	wanted	actions	skill
devour quickly	use sparingly	unpleasant work	timing
necessity	work	female skill	fulfilling
wanted			

Experiment 1 and extracted the alignment property (i.e. the way in which the participant identified the 2 concept pairs to be similar). For instance, in the response “money and chocolate are alike because they both make you happy”, the alignment property was identified as “happy” (see Table 2 for the alignment property stimuli identified by the coders). Coder consensus was 100%.

Generation of output dominance lists for individual concepts

A separate group of 28 college undergraduates (14 men and 14 women) were asked to generate a list of 20 properties for each of the 8 concepts used from Experiment 1. A set of three independent coders then sorted the features for each concept in terms of output dominance (i.e. frequency of occurrence) across the 28 participants; this resulted in 8 separate output dominance-ordered lists, 1 per concept.

Phase 2: determination of output dominance of features

Participants and procedure

Participants for this phase of the study were 256 students in introductory psychology classes at a large southwestern university, who received credit toward class requirements by participating in this study. Participants were given a set of alignment properties drawn from the responses in Experiment 2, Phase 1 as well as the corresponding feature list for only one of the two constituent concepts; they did not see the actual Funny or Not Funny responses from which the alignment properties had been drawn. Each set of alignment properties and feature lists was assessed by 16 participants, who were to find any matches between the alignment properties and features. For example, eight participants received the alignment property (e.g. “hot water”) drawn from responses to the concept pair “cooking and cheating” and were to match these with the features listed for “cooking”, whereas a counterpart eight participants were given the same set of alignment properties but were to match them with the feature list for “cheating”.

The alignment properties (refer to Table 2) were presented in two different fixed random orders, and no participant received more than five alignment properties and one concept feature list. A letter (e.g. A, B, C, etc.) was presented next to each alignment

property, and participants were to write down the appropriate letter next to any of the listed features they considered to be synonymous with that alignment property. Participants were informed that it was allowable to match more than one letter (or no letters) with a feature according to their own perceptions of related meaning. Participants were told to take as much time as needed, and all completed the task within 20 minutes.

Importantly, having the larger sample of participants in Experiment 2 judge the synonymy of the features generated by the sample of 28 participants in Phase 1 above and the alignment properties identified by the coders allowed a degree of validation of both as measures of semantic content. For instance, the alignment property “lack of” (which coders drew from a participant’s response to the juxtaposition of “money and sex”) was independently judged by a separate Experiment 2 participant to be semantically equivalent to the feature “deficit”.

Once participants’ judgments of synonymy between alignment properties and feature lists were collected, output dominance scores were computed for each alignment property a given participant received (their subset from the original 16 cells given in Table 2). The *alignment property* output dominance scores were assigned by summing the output dominance scores (frequency of mention) of all the *features* the participant marked as synonymous to that alignment property (recall that the output dominance scores for the features—but not the alignment properties—had been produced by separate participants in Experiment 1). For instance, an Experiment 2 participant identified the alignment property “work” as synonymous with the features “employment” and “hard work” under the concept “money”. Each of these features had produced an output dominance score of 2 in Experiment 1, so these scores were summed to produce an output dominance score of 4 for the alignment property “work” for that the Experiment 2 participant. Final output dominance scores were produced for each alignment property by taking the mean of scores produced by Experiment 2 participants who received the same alignment property and the same concept feature list. Responses from 35 participants were rejected, as these were either incomplete or did not follow task instructions, leaving 221 viable responses.

Because each alignment property was presented to Experiment 2 participants with either one or the other

of the paired concepts from Experiment 1, we were able to identify 2 independently produced output dominance scores for each alignment property in Experiment 2. For instance, the alignment property “hot water” produced one average output dominance score based on judgments from participants who scored the feature under the concept “cooking”, and another for participants who scored the same feature under the concept “cheating”.

Results

Recall our hypothesis was that significantly different output dominance scores would produce more instances of responses judged Funny. The two output dominance scores produced for each alignment property in Experiment 2 were entered into a *t*-test analysis for independent samples. Based on the observation that the relative variance of scores within cells was high, the data were subjected to the Cochran and Cox *p* approximation and Satterthwaite’s approximation for degrees of freedom, which showed that, for purposes of the null hypothesis, each pair of samples had equal variances. A detailed list of the means, standard deviations, and *t*-test scores for each of the alignment properties for the concept pairs tested is provided in the Appendix. A summary of the findings is provided below.

Money and chocolate

Results of the *t*-tests indicated that alignment properties for 29% of Funny responses had significantly different output dominance means ($p < .05$) for “money” versus “chocolate”, whereas none of the Not Funny responses had significantly different output dominance means for these two concepts.

Sex and money

For alignment properties drawn from Funny responses to the juxtaposition of “sex” and “money”, 13% had significantly different means ($p < .05$) and 25% were marginally different ($p < .10$). None of the Not Funny alignment properties were significant or marginally different in output dominance.

Cooking and cheating

Alignment properties drawn from Funny responses to the juxtaposition of “cooking” and “cheating” had the

highest frequency of significant differences in output dominance means. Thirty-eight percent achieved statistical significance at $p < .05$, and 25% were marginally significant ($p < .10$). For this concept pair, one of the alignment properties drawn from a response judged Not Funny also achieved statistical significance ($p < .05$), and one was marginally significant ($p < .10$).

Cooking and joking

Finally, for the juxtaposition of “cooking” and “joking”, 33% of Funny responses had significantly different output dominance means for the alignment properties ($p < .05$), but none were significantly different for Not Funny responses.

In summary, results indicated significantly different output dominance scores for 8 of the 29 alignment properties drawn from responses judged Funny, and an additional 4 were marginally different. In other words, *over 41% of Funny responses followed the predicted pattern*. In contrast, only one alignment property drawn from Not Funny responses showed significantly different output dominance scores, and one more was marginally different. That is, *over 97% of Not Funny responses followed the predicted pattern* (see Figure 4).

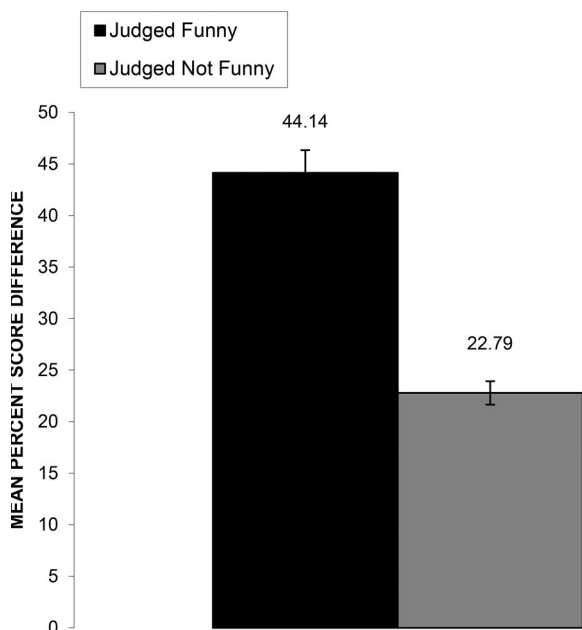


Figure 4. Mean percent divergence in concept pair output dominance scores for responses judged “Funny” vs. “Not Funny” in Experiment 2.

The relative treatment magnitude was calculated for each comparison that yielded a significant *t*. All effect sizes (partial eta squares) fell between 0.09 and 0.22. Based on Cohen's labels for use in the behavioural sciences (see Keppel & Wickens, 2004), these effect sizes are in the "medium" to "large" range, accounting for approximately 10–20% of the considerable variance in the raw scores.

In summary, our findings provided support for our predictions: compared to responses judged not funny, responses judged funny were much more likely to have a significant mismatch in output dominance of the alignment properties associated with the individual concepts involved. This result supports the notion that the alignment of features with incongruous output dominance provides a useful way of predicting what people will find funny.

General discussion

Across two experiments, the goal of this research was to examine the contribution of different factors to the production of verbal humour and to identify a cognitive mechanism underlying humorous production of concept comparisons. The factors examined were the context in which humour is elicited (incidental vs. intentional), the humour producer's gender, the affective content of the humour-eliciting stimuli, whether the stimuli were related or unrelated in meaning, and whether the stimuli were to be compared for similarities or differences.

The results from Experiment 1 showed that humour responses were more evident under incidental than intentional humour production instructions. This finding is consistent with an outcome noted in creativity studies whereby instructing people to "be creative" is not an effective way of eliciting creative responses. Our study provides the first demonstration that instructing people to "be funny" is not as effective as asking them to "be catchy" in how they frame their responses. One interpretation of this effect is that being asked to be funny leads to a kind of performance anxiety, since generating humour has clear performance aspects. Another interpretation is that humour production for the kind of task we used was cognitively demanding.

With respect to the claim that men are more likely to initiate humour than women, our study found no support. Instead, we found an interaction of gender and instructional set. Both men and women produced more humorous responses when the instructions were

to be catchy than to be funny, but this effect was heightened in men. This effect may hint at a gender difference in performance-related anxiety, with men performing less well when asked to be funny than when not asked to be funny. That is, men's humorous responses were heightened when there was less overt pressure to be funny. However, their percent of humorous responses in the "be catchy" condition still did not exceed that of women. As such, our study offers no support for the fitness-based claim that men are more likely than women to produce humour.

With regard to the role of affective content of responses, we had predicted that humorous responses would be more likely when the content includes a socially inappropriate topic (e.g. sex or cheating) than when the topic is not affectively charged in this way. Although we found no evidence for an overall effect of affective salience, there was a trend towards an effect in interaction with instructional set: under incidental humour production conditions, so-called taboo items tended to show a higher level of response than non-taboo items. Nevertheless, this was not a robust effect. The lack of a clear effect of affective salience may reflect the (relatively conservative) nature of our participant sample or it may suggest that our manipulation was not sufficiently sensitive. Future work will need to address this variable in other ways.

Finally, our study provided strong support that incongruity enhances humour generation. Construed as the juxtaposition of two or more dissimilar concepts, schemata, or meanings, incongruity has been considered a central mechanism underlying humour comprehension (Suls, 1972). In Experiment 1 the working of incongruity was evident in the finding that a greater number of humorous responses were produced when the concepts to be paired were unrelated rather than related in meaning, and when the task was to find differences rather than similarities between the concepts.

Experiment 2 further probed the working of incongruity. We had predicted that humour would arise from aligning incongruous aspects of two concepts. We tested the role of incongruity in humour generation by examining whether significantly divergent output dominance scores of juxtaposed concept properties would predict the use of those properties in verbal productions judged funny. We found this to be the case. The results supported a consistent directional relationship between humour content and

degree of output dominance divergence between the base and target features involved, such that the base concept contains a feature of high output dominance, while the same feature in the target has a low dominance. Overall, the results indicated that alignment properties drawn from Funny responses had significantly different default output dominance scores relative to Not Funny responses in each of the paired concepts, as predicted.

One limitation of the present study was that we did not code the responses in terms of degree of humorousness. As such, it is conceivable that those Funny alignment properties that did not have significantly different output dominance scores may have been drawn from productions most participants found only slightly amusing or perhaps from productions found funny by our judges, but not by a larger sample of participants. These issues should be more systematically addressed in future experiments.

Another limitation is that there were only six to eight participants in each of the cells in Experiment 2. Using a larger sample may have yielded a more robust pattern of results. Furthermore, the use of a free production task introduces a high degree of variation, more so than might be expected in a more constrained perception paradigm (Blanchette & Dunbar, 2000). Taken together, these two facts may have contributed to a relatively lower power for the study. Further examination of factors such as the high degree of natural heterogeneity in the impressions about what features are similar are needed in future studies, perhaps through the collection of normative data on synonyms for alignment properties prior to analysis.

Despite these limitations, the finding of significant differences in output dominance scores in a sizeable percentage of the Funny responses, together with the marked absence of significant differences in the Not Funny responses, argues in support of the research hypothesis tested. That is, incongruity, as indexed by divergence of output dominance in constituent concept features, is more likely to occur in humorous than non-humorous verbal productions.

With respect to the Setup-Incongruity-Resolution (SIR) theoretical model of humour processing (Attardo, 1997), the present study found support for both the setup and the incongruity stages. The setup phase directly corresponds in the present research to the instantiation of default properties for concepts activated prior to the humorous juxtaposition (e.g. "fattening" was very seldom deemed a feature of "money" when "money" was presented in isolation, but

participants immediately accepted the validity of "fattening" as a feature when "money" was juxtaposed with "chocolate"). The present research also found support for Attardo's (1997) incongruity stage of humour processing in that significantly more alignment properties drawn from Funny responses involved divergent output dominance scores (operationalised as incongruity) relative to alignment properties drawn from Not Funny responses. The resolution stage of Attardo's (1997) model of humour processing has yet to be empirically tested. Future research might incorporate a methodology similar to that used in the present research to test the resolution stage of IR models, for example, by assessing any changes in the default output dominance scores of alignment properties in concepts presented separately, but, in this case, after concept juxtaposition as opposed to before.

The present research was not designed to test the relative validity of different models of conceptual combination; rather, the research sought to identify models valid in other domains that could also persuasively describe the data generated in the domain of humour. Structural alignment (Markman & Gentner, 1993, 2000), as one such model, may address the changing salience of features during the processing of humorous juxtapositions as follows: (1) the humour producer draws the hearer's attention to some surface similarity or relational structure (i.e. the alignment property) between the initial expectations and the incongruous ones; (2) the two concepts are structurally aligned, and the humour perceiver may then draw inferences from an alignment property's role in the base (favoured) concept to that of its counterpart in the target (surprising) concept. When such inferences involve a significant increase in a feature's salience (output dominance) in the target schema, humour is likely to be found.

Another model that may be useful in describing our data is that of concept specialisation (Murphy, 1997), which also relies on dynamic features. In the context of concept comparisons, concept specialisation allows the substitution of features from one concept with those from another, and the tailoring of concepts using world knowledge.

A third model, frame blending (Coulson, 2001), may be used to account for the present data as well. In this model, two or more frames (or concepts) are activated in parallel, and mappings between the concepts allow integration and transfer of features, accounting for the changing salience of features during humour processing. Although the structural alignment model, the

concept specialisation model, and the frame blending model were not formally tested in the present study, the data are adequately described by analogous operations represented in each. Specific testing of these models in a humour domain would be useful both for describing aspects of humour processing not currently addressed and for extending the scope and strength of each model.

It is important to note that each of the three cognitive models of concept comparison described in this study relies on a schema-based structure of mental representation (as opposed to a semantic network, for instance). Schema models appear to be particularly well suited to describing mental representations of concepts as clusters of knowledge with features that can be reorganised as needed to interpret a given experience (as in humour). Given that the perception of humour relies on a change in conceptual understanding, dynamic aspects of these three models may provide useful ways of theorising about cognitive processes underlying humour creation.

Finally, it should be pointed out that existing cognitive models such as those described above will at some point need to incorporate findings from emotion-based accounts of humour functioning in social interaction to arrive at a richer depiction of how humour may operate in the service of emotion regulation (e.g. Samson, 2012). These emotion-based accounts suggest that humour serves a variety of functions related to the management of human emotions, ranging from providing empathy, solace, comfort, diversion, or cognitive reappraisal of negative emotion-eliciting events or experiences.

Conclusion

Our study provides an empirical grounding for the notion of incongruity typically invoked in cognitively based models of humour processing. In addition to providing an independent method for predicting what comparisons people will tend to find funny, the study addresses where and how in verbal humour processing incongruity occurs in a concept juxtaposition context. Specifically, we propose that incongruity occurs at the feature level by projecting the salience of a high output dominant feature in one concept to another concept in which the feature has a low default salience.

Results of this study suggest that output dominance scores may serve as a mechanism to quantify the cognitive distance between features of juxtaposed concepts, which, in turn, provides for an operationalisation of

incongruity in terms of a measurable difference in default output dominance between aligned properties in the base and target of a humorous verbal production. Our study thus offers a means of quantifying the condition of “surprisingness” many humour theorists agree is necessary (if not sufficient) to produce humour.

Note

1. The Holm-Bonferroni correction was used to minimize the possible inflation of the Type I error risk due to multiple paired comparisons. p' scores demonstrated corrected p scores.

References

- Aarons, D. (2012). *Jokes and the linguistic mind*. New York: Routledge.
- Aillaud, M., & Piolat, A. (2012). Influence of gender and judgment of dark and nondark humor. *Individual Differences Research*, 10(4), 211–222.
- Attardo, S. (1997). The semantic foundations of cognitive theories of humor. *Humor*, 10, 395–420.
- Attardo, S., & Raskin, V. (1991). Script theory revis(it)ed: Joke similarity and joke representation model. *Humor: International Journal of Humor Research*, 4(3/4), 293–347.
- Barsalou, L. W., & Prinz, J. J. (1997). Mundane creativity in perceptual symbol systems. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought* (pp. 267–307). Washington, DC: American Psychological Association.
- Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory & Cognition*, 28, 108–124.
- Canestrani, C., & Bianchi, I. (2012). Perception of contrariety in jokes. *Discourse Processes*, 49, 539–564.
- Chafe, W. (2007). *On the importance of not being earnest*. Amsterdam: John Benjamins.
- Coates, J. (2014). Gender and humor in everyday conversation. In D. Chiaro & R. Baccolini (Eds.), *Gender and humor: Interdisciplinary and international perspectives* (pp. 145–166). New York: Routledge.
- Coulson, S. (2001). *Semantic leaps: Frame-shifting and conceptual blending in meaning construction*. Cambridge: Cambridge University Press.
- Coulson, S., Urbach, T. P., & Kutas, M. (2006). Looking back: Joke comprehension and the space structuring model. *Humor: International Journal of Humor Research*, 19, 229–250.
- Cunningham, W., & Derks, P. (2005). Humor appreciation and latency of comprehension. *Humor: International Journal of Humor Research*, 18(4), 389–403.
- Douglas, M. (1975/1991). Jokes. In C. Mukerji & M. Schudson (Eds.), *Rethinking popular culture: Contemporary perspectives in cultural studies* (pp. 291–310). Berkeley, CA: University of California Press.
- Du, X., Qin, Y., Tu, S., Yin, H., Wang, T., Yu, C., & Qiu, J. (2013). Differentiation of stages in joke comprehension: Evidence from an ERP study. *International Journal of Psychology*, 48, 149–157.

- Freud, S. (1905/1960). *Jokes and their relation to the unconscious*. New York: Norton. (Original work published 1905.)
- Giora, R. (1991). On the cognitive aspects of the joke. *Journal of Pragmatics*, 16, 465–485.
- Giora, R., Fein, O., Kronrod, A., Elnatan, I., Shuval, N., & Zur, A. (2004). Weapons of mass distraction: Optimal innovations and pleasure ratings. *Metaphor and Symbol*, 19(2), 115–141.
- Greengross, G., & Miller, G. (2011). Humor ability predicts intelligence, predicts mating success, and is higher in males. *Intelligence*, 39, 188–192.
- Gueguen, N. (2010). Men's sense of humor and women's responses to courtship solicitations: An experimental field study. *Psychological Reports*, 107, 145–156.
- Hempelmann, C. F., & Ruch, W. (2005). 3 WD meets GTVH: Breaking the ground for interdisciplinary humor research. *HUMOR: International Journal of Humor Research*, 18(4), 353–387.
- Hillson, T. R. & Martin, R. A. (1994). What's so funny about that? The domains-interaction approach as a model of incongruity and resolution in humor. *Motivation and Emotion*, 18(1), 1–29.
- Kana, R. K., & Wadsworth, H. M. (2012). "The archeologist's career ended in ruins": Hemispheric differences in pun comprehension in autism. *NeuroImage*, 62, 77–86.
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Koestler, A. (1964). *The act of creation*. New York: MacMillan.
- Kozbelt, A., & Nishioka, K. (2010). Humor comprehension, humor production, and insight: An exploratory study. *Humor*, 23(3), 375–401.
- Kugler, L., & Kuhbandner, C. (2015). That's not funny! – But it should be: Effects of humorous emotion regulation on emotional experience and memory. *Frontiers in Psychology*, 6, Article No. 1296, 1–7. doi:10.3389/fpsyg.2015.01296
- Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, 32, 517–535.
- Markman, A. B., & Gentner, D. (2000). Structure-mapping in the comparison process. *The American Journal of Psychology*, 113, 501–538.
- Martin, R.A. (2014). Humor and gender: An overview of psychological research. In D. Chiaro & R. Baccolini (Eds.), *Gender and humor: Interdisciplinary and international perspectives* (pp. 123–144). New York: Routledge.
- Mickes, L., Walker, D.E., Parris, J.L., Mankoff, R., & Christenfeld, N.J.S. (2012). Who's funny: Gender stereotypes, humor production and memory bias. *Psychonomic Bulletin & Review*, 19, 108–112.
- Miller, G. F. (2000). Mental traits as fitness indicators: Expanding evolutionary psychology's adaptationism. In D. LeCroy & P. Moller (Eds.), *Evolutionary perspectives on human reproductive behavior*, *Annals of the New York Academy of Sciences, Volume 907* (pp. 62–74). Baltimore, MD: Johns Hopkins University Press.
- Minsky, M. (1984). Jokes and the logic of the cognitive unconscious. In L. Vaina & J. Hintikka (Eds.), *Cognitive constraints on communication* (pp. 175–200). Dordrecht: D. Reidel.
- Murphy, G. (1997). Polysemy and novel word meanings. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought* (pp. 235–265). Washington, DC: American Psychological Association.
- Niu, W., & Sternberg, R.J. (2001). Cultural influences on artistic creativity and its evaluation. *International Journal of Psychology*, 36, 225–241.
- Norrick, N. R. (1986). A frame-theoretical analysis of verbal humor: Bisociation as schema conflict. *Semiotica*, 60, 225–245.
- Ruch, W. (1992). Assessment of appreciation of humor: Studies with the 3WD humor test. In C. D. Spielberger, & J. N. Butcher (Eds.), *Advances in personality assessment* (Vol. 9, pp. 27–75). Hillsdale, NJ: Erlbaum.
- Runco, M., Illies, J., & Reiter-Palmon, R. (2005). Explicit instructions to be creative and original: A comparison of strategies and criteria as targets with three types of divergent thinking tests. *The Korean Journal of Thinking & Problem Solving*, 15(1), 5–15.
- Samson, A. C. (2012). The influence of empathizing and systemizing on humor processing: Theory of mind and humor. *Humor*, 25, 75–98.
- Samson, A., Glassco, A., Lee, I., & Gross, J. (2014). Humorous coping and serious reappraisal: Short-term and longer-term effects. *Europe's Journal of Psychology*, 10(3), 571–581.
- Samson, A. C., Hempelmann, C. F., Huber, O., & Zysset, S. (2009). Neural substrates of incongruity-resolution and nonsense humor. *Neuropsychologia*, 47, 1023–1033.
- Schmidt, S. (2002). The humour effect: Differential processing and privileged retrieval. *Memory*, 10(2), 127–138.
- Strick, M., Holland, R., Van Baaren, R., & Knippenberg, A.V. (2010). Humor in the eye tracker: Attention capture and distraction from context cues. *The Journal of General Psychology*, 137(1), 37–48.
- Suls, J. M. (1972). A two-stage model for the appreciation of jokes and cartoons. In J. Goldstein & P. McGhee (Eds.), *Psychology of humor* (pp. 81–100). New York: Academic Press.
- Vaid, J. (1999). The evolution of humor: Do those who laugh last? In D. H. Rosen & M. Luebbert (Eds.), *Evolution of the psyche* (pp. 123–138). Westport, CT: Praeger.
- Vaid, J. (2006). Joking across languages: Perspectives on humor, emotion, and bilingualism. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression and representation* (pp. 152–182). Clevedon: Multilingual Matters.
- Vaid, J. (2014). 'Absence makes the heart grow fonder ... and also makes the eyes wander'. Optimal innovations in proverb rejoinders. In A. Kozbelt (Ed.), *Proceedings of the international conference on empirical aesthetics* (pp. 1–4). New York City: Hunter College.
- Vaid, J., Hull, M.R., Heredia, R., Gerkens, D., & Martinez, F. (2003). Getting a joke: The time course of meaning activation in verbal humor. *Journal of Pragmatics*, 35(9), 1431–1449.
- Vaid, J., & Kobler, J.B. (2000). Laughing matters: Toward a structural and neural account. *Brain and Cognition*, 42, 139–141.
- Veale, T. (2008). Figure-ground duality in humour: A multi-modal perspective. *Lodz Papers in Pragmatics*, 4, 63–81.
- Veatch, T. C. (1998). A theory of humor. *Humor*, 11, 161–215.
- Wyer, R. S., & Collins, J. E. (1992). A theory of humor elicitation. *Psychological Review*, 99, 663–688.

Appendix

Table A1. Alignment property cell means, standard deviations and *t*-test results.

Concept pair with properties	Concept 1 mean	St. dev. 1	<i>t</i> -value	<i>p</i> -value	Concept 2 mean	St. dev. 2
MONEY & CHOCOLATE						
evil	17	17.95	-1.24	n.s.	29	20.31
sweet*	87	61.81	2.84	.0131	18	30.67
melt	29	25.28	-1.24	n.s.	22.75	24.38
negative outcome	17	14.81	0.72	n.s.	13	5.82
leaves hands quickly	10	13.31	-0.46	n.s.	14	20.12
stress relief/escape	16	16.6	0.03	n.s.	15	13.23
fattening* (attractive to women)	49	44.17	2.36	.0375	6	4.14
(devour quickly) (necessity)	26	30.46	1.09	n.s.	12	18.12
(wanted)	27	29.64	0.37	n.s.	22	21.98
(satisfying)	12	8.75	0.010	n.s.	12	11.57
(wanted)	23	20.77	0.08	n.s.	22	39.94
(easy to get)	9	9.52	-1.34	n.s.	28	30.12
(touchable)	8	3.08	-1.22	n.s.	19	19.73
SEX & MONEY	18	22.49	0.29	n.s.	14	25.33
awkward	13	19.54	0.38	n.s.	9	17.48
hard to get	13	12.4	-1.01	n.s.	28	35.43
diminish	14	12.24	-0.96	n.s.	22	20.48
never had	16	23.46	-0.042	n.s.	21	22.58
think about**	38	41.34	1.26	n.s.	18	22.43
politician	30	38.78	1.88	.0825	2	2.67
lack of*	4	6.16	0.91	n.s.	7	6.21
depletion**	1	2.50	-5.84	.0001	12	4.57
(wanted)	5	5.76	-1.87	.0843	10	5.96
(use sparingly)	62	78.37	0.66	n.s.	31	11.14
(work)	21	27.65	-0.017	n.s.	22	11.85
(bad)	13	14.11	-1.70	n.s.	33	26.56
(dirty)	28	33.28	-0.21	n.s.	32	32.20
(evil)	8	9.09	0.99	n.s.	4	4.05
(highly desired)	8	6.98	-0.80	n.s.	30	65.95
COOKING & CHEATING	56	39.32	0.40	n.s.	45	59.67
spicy	7	2.00	-0.82	n.s.	15	24.28
meat	15	11.27	0.57	n.s.	10	18.15
not fatal*	3	3.50	-2.62	.0256	29	24.06
hot water**	7	7.12	-2.01	.0722	22	17.06
clean up afterward	10	6.16	1.42	n.s.	5	5.15
shortcuts**	20	14.13	1.79	.0966	9	8.97
denial of*	36	31.98	2.62	.0212	7	7.99
hurt*	41	38.92	2.49	.0271	8	6.64
(give you gain)	27.67	16.64	0.88	n.s.	18	23.24
(actions)	49	41.85	0.83	n.s.	30	39.10

Table A1. Continued.

Concept pair with properties	Concept 1 mean	St. dev. 1	<i>t</i> -value	<i>p</i> -value	Concept 2 mean	St. dev. 2
(unpleasant work)	20	10.05	1.16	n.s.	13	11.27
(female skill)**	55	60.55	2.07	.0629	8	7.99
(practice)	38	48.21	0.50	n.s.	28	8.88
(skill)	30	34.67	0.75	n.s.	20	16.64
(bad outcome)*	34	22.46	2.56	.0252	11	12.35
COOKING & JOKING						
can make you sick	18	11.93	-0.09	n.s.	31	32.29
can cause discomfort	22	19.22		n.s.	37	16.48
laugh*	129	56.53	4.69	.0007	14	20.33
right ingredients*	10	8.76	-2.26	.0451	30	21.30
chicken	12	12.14	-1.21	n.s.	26	28.87
dad/male	19	21.55	0.50	n.s.	14	17.04
(spicy)	6	8.89	-1.01	n.s.	11	10.30
(wait)	7	7.35	-1.45	n.s.	15	11.90
(skill)	16	12.25	-0.04	n.s.	16	11.99
(timing)	15	17.27	0.20	n.s.	14	6.97
(fulfilling)	13	15.88	-1.55	n.s.	26	13.49
(skill/practice)	24	30.47	-0.31	n.s.	28	30.47
(produce enjoyment)	110	64.42	1.71	n.s.	60	51.81

Note: Alignment properties drawn from "Not Funny" responses are shown in parentheses. * $p < .05$; ** $p < .10$.

(Continued)